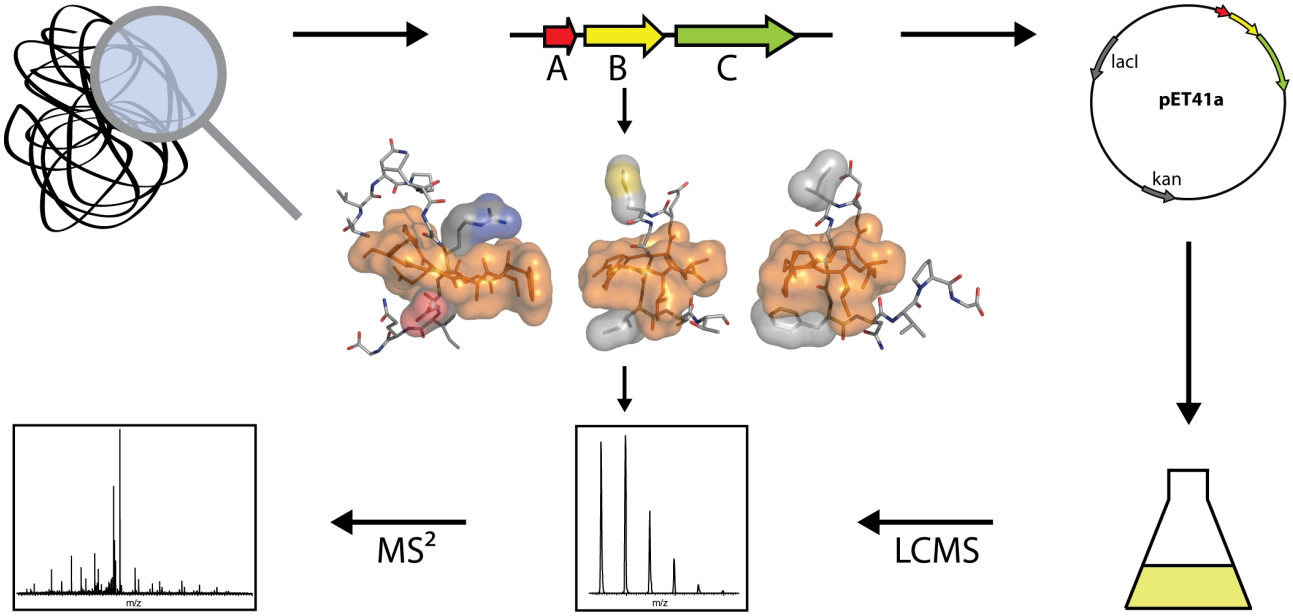
Genome Mining, Isolation and Characterization of Novel Lasso Peptides and Their Utilization in Drug Development

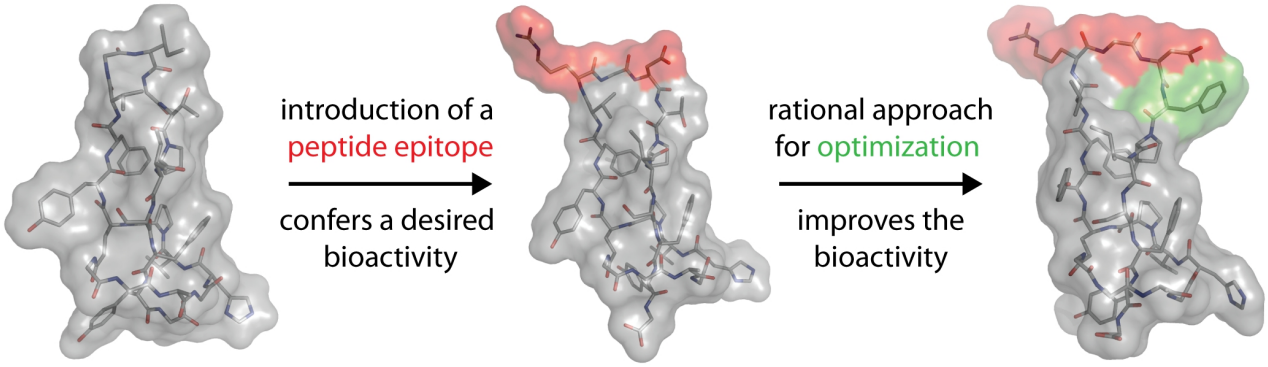
---

Genome Mining, Isolierung und Charakterisierung von neuartigen Lassopeptiden und ihre Nutzung in der Wirkstoffentwicklung

Dissertation

Julian Hegemann





Marburg an der Lahn, 2014

Genome Mining, Isolation and Characterization of Novel Lasso Peptides and Their Utilization in Drug Development

---

Genome Mining, Isolierung und Charakterisierung von neuartigen Lassopeptiden and Ihre Nutzung in der Wirkstoffentwicklung

Dissertation
zur Erlangung des Doktorgrades
der Naturwissenschaften
(Dr. rer. nat.)

dem
Fachbereich Chemie
der Philipps-Universität Marburg
vorgelegt von

Dipl.-Chem. Julian David Hegemann
aus Marburg an der Lahn (Hessen)

Marburg an der Lahn, 2014

Vom Fachbereich Chemie
der Philipps-Universität Marburg (Hochschulkennsziffer 1180)
als Dissertation am _____ angenommen.

Erstgutachter:	Prof. Dr. Mohamed A. Marahiel
	(Fachbereich Chemie, Philipps-Universität Marburg)
Zweitgutachter:	Prof. Dr. Peter Graumann
	(Fachbereich Chemie, Philipps-Universität Marburg)

Tag der Disputation: 14.08.2014

Dedicated to my friends and family

The majority of the work presented herein has been published:

Julian D. Hegemann, Marcel Zimmermann, Xiulan Xie, and Mohamed A. Marahiel, *Caulosegnins I-III: A Highly Diverse Group of Lasso Peptides Derived From a Single Biosynthetic Gene Cluster*, *J. Am. Chem. Soc.* **2013**, *135(1)*, 210-222.

Marcel Zimmermann, **Julian D. Hegemann**, Xiulan Xie, and Mohamed A. Marahiel, *The Astexin-1 Lasso Peptides: Biosynthesis, Stability, and Structural Studies*, *Chem. Biol.* **2013**, *20(4)*, 558-569.

Julian D. Hegemann, Marcel Zimmermann, Shaozhou Zhu, Dennis Klug, Mohamed A. Marahiel, *Lasso Peptides From Proteobacteria: Genome Mining Employing Heterologous Expression and Mass Spectrometry*, *Biopolymers Pept. Sci.* **2013**, *100(5)*, 527-542.

Julian D. Hegemann, Marcel Zimmermann, Shaozhou Zhu, Holger Steuber, Klaus Harms, Xiulan Xie, Mohamed A. Marahiel, Xanthomonins I-III: A New Class of Lasso Peptides With a Seven-Residue Macrolactam Ring, *Angew. Chem. Int. Ed. Engl.* **2014**, *53(8)*, 2230-2234 and *Angew. Chem.* **2014**, 126(8), 2262-2266.

Marcel Zimmermann, **Julian D. Hegemann**, Xiulan Xie, and Mohamed A. Marahiel, *Characterization of Caulonodin Lasso Peptides Revealed Unprecedented N-Terminal Residues and a Precursor Motif Essential for Peptide Maturation*, *Chem. Sci.* **2014**, *Epub ahead of print*. (doi: 10.1039/c4sc01428f)

Julian D. Hegemann, Mariarosaria De Simone, Marcel Zimmermann, Thomas A. Knappe, Xiulan Xie, Francesco Saverio Di Leva, Luciana Marinelli, Ettore Novellino, Stefan Zahler, Horst Kessler, and Mohamed A. Marahiel, *Rational Improvement of the Affinity and Selectivity of Integrin Binding of Grafted Lasso Peptides*, *J. Med. Chem.* **2014**, *57(13)*, 5829-5834.

Zusammenfassung

Lassopeptide sind eine Klasse von Naturstoffen, die der Familie der ribosomal-synthetisierten und posttranslational-modifizierten Peptide zugehörig sind. Sie sind gekennzeichnet durch das einzigartige Strukturmotiv des Lassoknotens, dessen Name sich von der Beobachtung ableitet, dass diese Topologie an den Knoten in einer Lassoschlinge erinnert. Diese Struktur wird durch einen *N*-terminalen Makrolactamring verwirklicht, der von dem *C*-terminalen Peptidschwanz durchfädelt ist. Die Faltung dieser Moleküle wird durch nicht-kovalente Wechselwirkungen sterisch anspruchsvoller Aminosäuren, die über und unter dem Ring lokalisiert sind, aufrecht erhalten, wodurch der *C*-terminale Bereich des Peptids im Inneren des Ringes fixiert wird. Was diese Verbindungen interessant für die Forschung macht ist, dass ihre Struktur trotz ihrer rein sterischen Stabilisierung oft eine außergewöhnliche Resistenz gegenüber thermischer, chemischer und proteolytischer Degradation aufweist. Dennoch ist bisher wenig über die allgemeine Funktion dieser Stoffe in der Natur bekannt, obwohl es einige zuvor beschriebene Lassopeptide gibt, für die interessante biologische Aktivitäten beobachten wurden.

Um mehr Information über ihre physiko-chemischen Eigenschaften, ihre Biosynthese und ihre mögliche Rolle in der Natur zu erhalten, war das primäre Ziel dieser Arbeit das gezielte *Genome Mining* und die darauffolgende Isolierung und Charakterisierung neuartiger Lassopeptide. Die Ergebnisse dieser Untersuchungen wurden in mehreren Studien publiziert, die im Verlauf dieser Doktorarbeit präsentiert und diskutiert werden sollen. Neben der Isolierung einer Vielzahl neuer Lassopeptide, konnten durch die sorgfältige Untersuchung und Charakterisierung dieser neuen Verbindungen zusätzlich einige bisher vorherrschende Vorstellungen aus diesem Forschungsgebiet widerlegt und aktualisiert werden. Ferner hat die Auswertung der Daten aus unseren *Genome-Mining*-Studien interessante Beobachtungen über das Vorkommen von Lassopeptiden in Bakterien und über die Existenz verschiedener Subgruppen von Biosynthesegencluster-andordnungen zugelassen. Diese Informationen könnten zukünftige Untersuchungen erleichtern, die auf die Identifizierung der konkreten Funktionen dieser Verbindung ausgelegt sind.

Des Weiteren wurde untersucht ob Lassopeptide geeignete Gerüste für die Wirkstoffentwicklung durch *Epitope-Grafting*-Ansätze sind. In dieser Hinsicht wurde eine vorher veröffentlichte, bioaktive Lassopeptidvariante als Basis genommen, um zu zeigen, dass solche Verbindungen in der Tat durch rationale Ansätze optimiert und verbessert werden können. Dieses Vorgehen basierte auf Forschungsergebnissen, die aus Studien mit linearen und zyklischen Peptiden erhalten wurden, welche, im Gegensatz zu Lassopeptiden, leicht durch chemische Synthese zugänglich sind.

Summary

Lasso peptides are a class of natural products that belong to the family of ribosomally-assembled and posttranslationally-modified peptides. They are defined by an unique structural motif referred to as the so-called lariat knot, whose name is derived from the fact that this topology is reminiscent of the knot found in the noose of a lasso. This structure is achieved by the presence of an *N*-terminal macrolactam ring that is threaded by the *C*-terminal tail of the peptide. The fold in these molecules is then conserved by non-covalent interactions in the form of bulky amino acids located above and below the macrolactam ring, in this way entrapping the tail inside of the ring. What makes these compounds of interest for research is that their structure, even though it is maintained merely by sterical interactions, often exhibits a tremendous stability against thermal, chemical and proteolytic degradation. Still, up to now little is known about the general function of these compounds for their producing organisms, although there are some interesting biological activities attributed to some of the previously reported lasso peptides.

To obtain more information about their physico-chemical properties, their biosynthesis and to get an idea what role they might play in nature, the primary subject of this thesis was the directed genome mining for and the subsequent isolation and characterization of novel lasso peptides. The results of these projects were published in several studies that will be shown and discussed in the course of this thesis. Amongst other findings, these studies not only include the discovery of a multitude of novel lasso peptides, but through the thorough analysis and characterization of these compounds, several former assumptions of this research area could be overhauled and updated. In addition to this, the bioinformatic data gathered during our genome mining studies furthermore uncovered interesting facts about the distribution of lasso peptides amongst bacteria and about the existence of different subgroups of biosynthetic gene cluster arrangements, which could facilitate future research directed towards identifiying the concrete functions of these compounds.

Furthermore, it was also investigated if these compounds are suitable scaffolds for drug development via epitope grafting approaches. In this regard, a previously reported bioactive lasso graft was used as the basis to show that such compounds can indeed be further optimized and improved upon by rational approaches that utilize the information obtained from research done with simple linear or cyclic peptides that are, in contrast to lasso peptides, easily accessible by synthetic means.

Table of Contents

1. Introduction

1.1 Ribosomally Produced Peptides

Natural products can originate from many sources both in regard to the producing organisms as well as to the way they are synthesized. A lot of well known examples of natural products belong to the groups of non-ribosomal peptides, polyketides, alkaloids, terpenoids or cyclic dipeptides. All of these compounds have in common that they are synthesized by enzymes utilizing the chemical building blocks available to the producing organisms. Therefore, alterations in these scaffolds by biosynthetic means can be difficult to introduce and depend on the availability of the building blocks and chemical variations thereof as well as on the ability of the corresponding enzymes to recognize and modify slightly different substrates.[1] In contrast to these groups of natural products, there is the class of so-called ribosomally-synthesized and posttranslationally-modified peptides (RiPPs), which utilizes gene encoded precursor peptides in combination with processing enzymes that transform them into diverse compounds with interesting biological properties.[2] Even though their genetic origin limits the precursor peptides to contain only proteinogenic amino acids, posttranslational modifications can introduce such dramatic changes to these peptide scaffolds that the resulting products may only barely resemble anything that was originally synthesized by the ribosome.[3-8] Nevertheless, although drastic changes may occur during the processing of a precursor peptide, some classes of RiPPs are also maintaining the majority of their original amino acids unaltered and introduce only slight changes that primarily alter the three-dimensional structure of the peptide scaffold and through these rather simple means generate a desired functionality and a high compound stability. Interestingly, there are also examples of RiPPs that in their mature form are not even peptides anymore, but only mere small organic molecules like pyrroloquinoline quinone (PQQ), even though their synthesis still starts with the ribosomal assembly of a precursor peptide.[3-5,7] Typically, all precursor peptides contain a core peptide that gets modified during maturation and is usually flanked by a sequence on at least one side that is needed for enyzmatic recognition, although it will not be present in the final natural product (Figure 1.1).[2]

In this work, the focus was directed upon representatives of the still young RiPP family of lasso peptides. Even though they have little complexity on a chemical level, they are a very interesting family of natural products as they feature unique and so far synthetically completely inaccessible three-dimensional structures combined with high sequence and functional diversity.

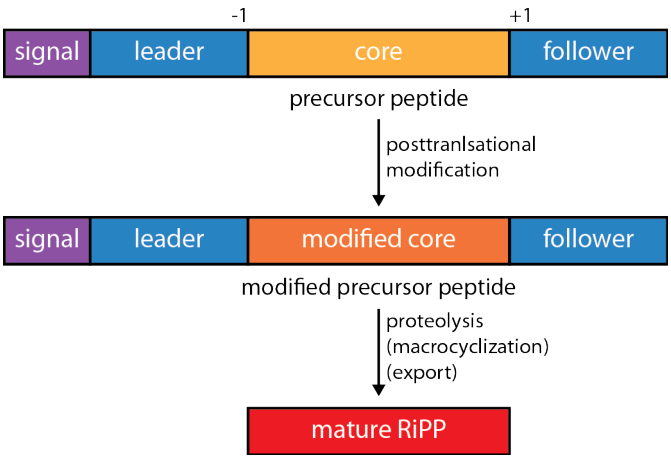


Figure 1.1. Schematic depiction of RiPP maturation. The different regions of the peptide during the different steps of the maturation are represented by colored boxes. Not every RiPP precursor features all of these regions, but may indeed contain only some of them, e.g. only a leader peptide and core peptide region. In general, only the core peptide gets posttranslationally modified. Most modifications are introduced while the precursor peptide is still intact, but final modification reactions, like macrocyclizations, may require prior proteolysis of the precursor peptide. Macrocyclization reactions and/or subsequent export of the mature compound is known to occur for some RiPPs, but is not observed for all of these natural products.

1.2 Lasso Peptides

1.2.1 Characteristics of Lasso Peptides

The RiPP family of lasso peptides is predominantly defined by their unique structural motif, the so-called lasso fold. This name was coined by the observation that the three-dimensional appearance of lasso peptides is reminiscent of a lariat knot (Figure 1.2).[9-30]

In general, the lasso fold is defined by the presence of a macrolactam ring through which the C-terminal tail is threaded. The tail is kept in place by two bulky residues, the so-called plugs, positioned above and below the macrolactam ring, stabilizing the structure merely by non-covalent interactions. Nevertheless, this steric fold conservation confers tremendous stability against chemical, proteolytical as well as in some cases also thermal degradation.[9,10,15,17,22,25-27,31-39] The macrolactam ring itself is generated by the formation of an isopeptide bond between the *N*-terminal α-amino group and a carboxylic acid side chain of an aspartate or glutamate residue that can be located at positions 7 to 9 of the lasso peptide. Generally, lasso peptides can be divided into three classes, which are defined by the number of disulfide bonds present. Class I lasso peptides have two disulfide bonds, of which one is formed with the for this class conserved cysteine residue at

position 1. Class II lasso peptides, which currently comprise the largest number of known lasso peptides, feature no disulfide bonds and therefore rely only on the non-covalent fold conservation conferred by the plug amino acids. The third and youngest class of lasso peptides contains a single disulfide bond and currently has only one known representative. An overview over all known lasso peptides, their primary structures and their class designation can be found in Table 1.1.

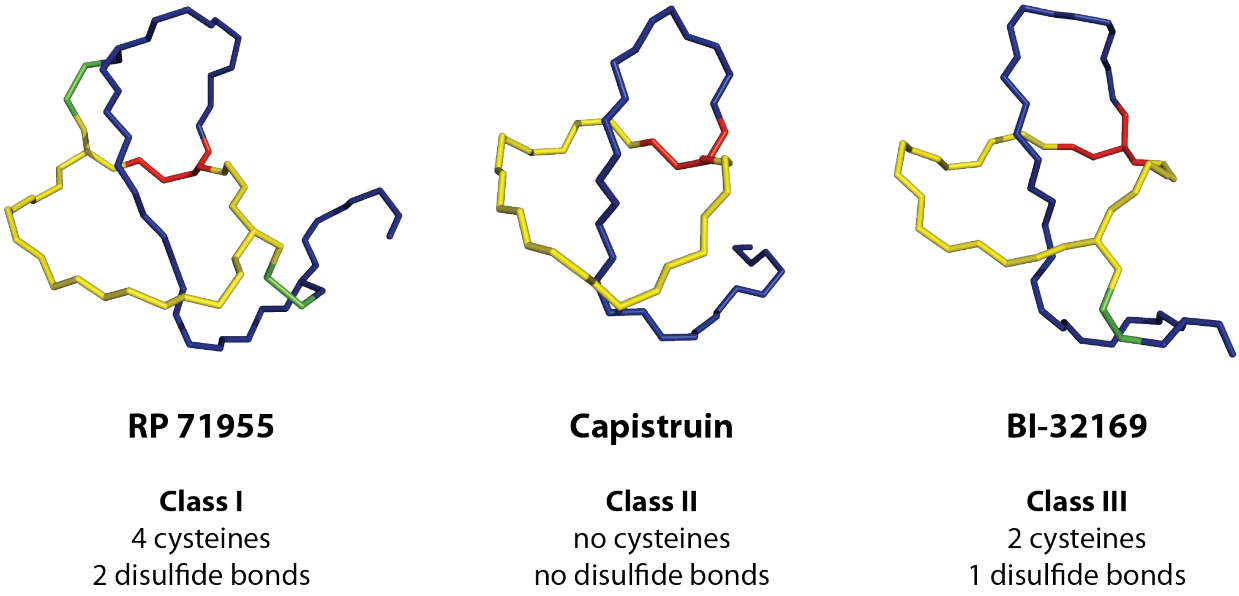
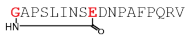
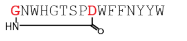
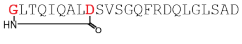
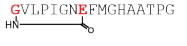
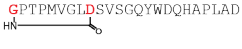
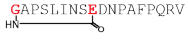
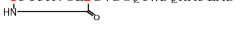
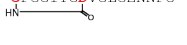
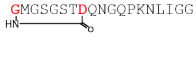
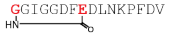


Figure 1.2. Backbones of representative lasso peptides belonging to the three different classes known for this group of natural products. The ring forming residues are shown in red, the remaining residues of the macrolactam in yellow and the amino acids of the tail region in blue. The disulfide bonds between two cysteine residues are highlighted in green. The shown lasso peptides are RP 71955 (PDB code 1RPB),[9] capistruin (BMRB code 20014)[18] and BI-32169 (PDB code 3NJW).[19,20]

The majority of three-dimensional structures of lasso peptides were solved via NMR spectroscopy,[9-19,21-26,28,30,49,50] while only two lasso peptide crystal structures were published so far.[20,26] Nonetheless, the NMR based structure elucidation of a lasso peptide is far from trivial.[21] This may explain how both the structures of microcin J25 (MccJ25)[50-53] and BI-32169[48] were first wrongly reported as head-to-tail cyclized or dicyclic peptides, respectively. Additionally, the recently published non-threaded structure of lassomycin[28] is highly debatable in regard to the facts that a branched-cyclic peptide should be rather susceptible towards proteolytic degradation and that such a defined solution structure is very unlikely for a compound with such a mobile *C*-terminal tail. Furthermore, it now appears that the first published solution structure of astexin-1[30] was also flawed, but in the opposite way and that indeed unthreaded astexin-1 was probed by NMR in this case. These assumptions are founded on subsequent experiments that showed that astexin-1 is

Table 1.1. Overview over all known lasso peptides with their host organisms, primary structures and class designations.

lasso peptide	source	primary structure	class	lasso peptide	source	primary structure	class
aborycin[10]	*Streptomyces griseoflavus* TÜ 4072		I	lariatin*[17]	*Rhodococcus jostii* K01-B0171		II
MS-271*[34]	*Streptomyces* sp. MS-271		I	lassomycin*[28]	*Lentzea kentuckyensis*		II
NP-06[27]	*Streptomyces* sp.		I	microcin J25*[32]	*Escherichia coli* AY25		II
RP71955*[40]	*Streptomyces* sp. SP9440		I	propetin[41,42]	*Microbispora* sp. SNA-115		II
siamycin I*[43]	*Streptomyces* sp. AA6532		I	RES-701-1[44]	*Streptomyces* sp. RE-701		II
siamycin II*[43]	*Streptomyces* sp. AA3891		I	RES-701-2[45]	*Streptomyces* sp. RE-701		II
SSV2083[29]	*Streptomyces sviceus* ATCC 20983		I	RES-701-3[45]	*Streptomyces* sp. RE-896		II
anantin[46]	*Streptomyces coerulescens*		II	RES-701-4[45]	*Streptomyces* sp. RE-701		II
astexin-1*[25,30]	*Asticcacaulis excentricus* CB48		II	rhodanodin[37]	*Rhodanobacter thiooxydans* LCS2		II
astexin-2*[23,25]	*Asticcacaulis excentricus* CB48		II	rubrivinodin[37]	*Rubrivivax gelatinosus* IL44		II
astexin-3[23]	*Asticcacaulis excentricus* CB48		II	sphingonodin I[37]	*Sphingobium japonicum* UT26		II
burhizin[37]	*Burkholderia rhizoxinica* HKI 454		II	sphingonodin II[37]	*Sphingobium japonicum* UT26		II
capistruin*[18]	*Burkholderia thailandensis* E264		II	sphingopyxin I[37]	*Sphingopyxis alaskensis* RB2256		II
caulonodin I[37]	*Caulobacter* sp. K31		II	sphingopyxin II[37]	*Sphingopyxis alaskensis* RB2256		II
caulonodin II[37]	*Caulobacter* sp. K31		II	SRO15-2005[29]	*Streptomyces roseosporus* NRRL 15998		II
caulonodin III[37]	*Caulobacter* sp. K31		II	sungsanpin[24]	*Streptomyces* sp.		II
caulonodin IV*[47]	*Caulobacter* sp. K31		II	syanodin I[37]	*Sphingobium yanoikuyae* XLDN2-5		II
caulonodin V[47]	*Caulobacter* sp. K31		II	xanthomonin I*[26]	*Xanthomonas gardneri* ATCC 19865		II
caulonodin VI[47]	*Caulobacter* sp. K31		II	xanthomonin II*[26]	*Xanthomonas gardneri* ATCC 19865		II
caulonodin VII[47]	*Caulobacter* sp. K31		II	xanthomonin III[26]	*Xanthomonas citri* pv. *magniferaeindicae* LM941		II
caulosegnin I*[22]	*Caulobacter segnis* ATCC21756		II	zucinodin[37]	*Phenylobacterium zucineum* HLK1		II
caulosegnin II[22]	*Caulobacter segnis* ATCC21756		II	BI-32169*[48]	*Streptomyces* sp. DSM 14996		III
caulosegnin III[22]	*Caulobacter segnis* ATCC21756		II				

7OHTrp stands for 7-hydroxy-tryptophan.

* The three-dimensional structure is reported for all lasso peptides marked with an asterisk.

thermally labile and that the plug amino acids suggested by the previous NMR structure were easily exchangeable with alanine. In contrast, a newly elucidated NMR structure of astexin-1 showed that the lower plug amino acid is indeed not R19 but Y15, which is in complete agreement with other observations that arose from a thorough mutational analysis.[25]

Concerning the origin and prevalence of lasso peptides, it is interesting to note that genome mining studies[30,54] revealed that putative lasso peptide biosynthetic gene clusters are broadly distributed amongst many bacterial phyla. Still, their actual role in nature remains elusive up to now.

1.2.2 Lasso Peptide Biosynthesis

Lasso peptide biosynthetic gene clusters consist of at least three genes encoding a precursor peptide A, a cysteine protease B and a macrocyclase C (Figure 1.3).[2] While in Gram-negative bacteria the majority of known and putative lasso peptide biosynthetic gene clusters have an ABC arrangement,[18,22,23,25,26,30,37,55,56] in Gram-positive bacteria an ACB formation is more prevalent,[28-30,57] suggesting a common predecessor that diverged in parallel with the development of Gram-negative and Gram-positive cell walls.

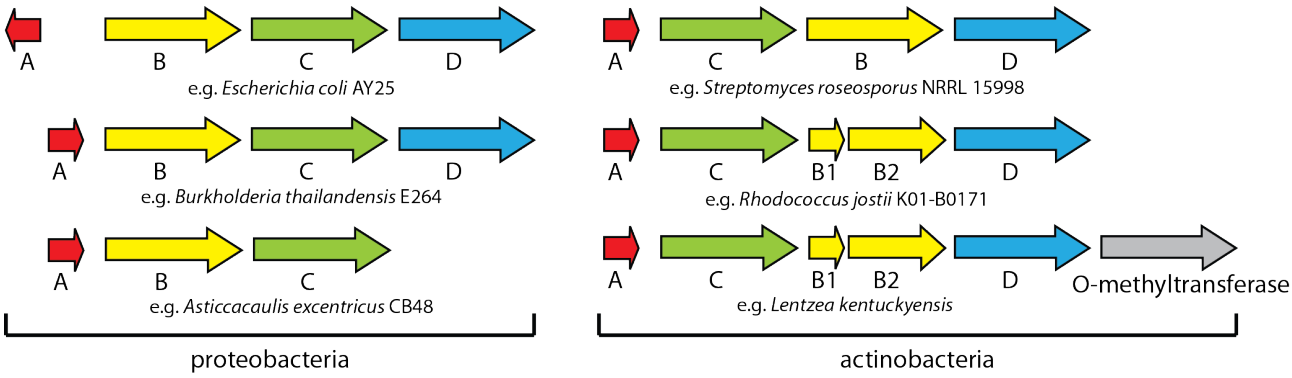


Figure 1.3. Schematic representation of the organization of known functional lasso peptide biosynthetic gene clusters with examples for host organisms they were found in. On the left side, organizations that are typical for proteobacterial lasso peptide biosynthetic gene clusters are shown. On the right side, typical organizations for actinobacterial lasso peptide biosynthetic gene clusters are depicted.

Other known lasso peptide biosynthetic gene clusters are complementing this three gene arrangement by a gene encoding an ATP binding cassette (ABC-) transporter D[18,28,29,37,57,58] and in one instance a cluster also featured a so far uncharacterized O-methyltransferase, which was proposed to be responsible for the methylated *C*-terminus of lassomycin that was produced homologously by fermentation of a *Lentzea* strain.[28] As previous genome mining studies revealed a number of lasso peptide biosynthetic gene clusters that comprised a variety of different enzymes so far unknown for these kind of clusters (amongst others kinases, sulfotransferases or

acyltransferases),[30,37] it seems reasonable to assume that these enyzmes could analogously modify their respective target lasso peptides. Indeed, the lasso peptides RES-701-2 and RES-701-4 are carrying another chemical modification in form of the hydroxylation of the aromatic moieties of their C-terminal tryptophan residues. As these two lasso peptides are otherwise identical with RES-701-1 and RES-701-3, respectively, this observation could suggest the presence of an enzyme in the producing strains that catalyzes the modification of tryptophan to 7-hydroxy-tryptophan for these lasso peptide scaffolds.[38,45] Still, because of the lack of genetic information this is merely speculative at the moment, even though such a selective chemical modification of these residues during the extraction and purification procedures appears to be rather unlikely. Another quite peculiar observation was made for the biosynthetic gene clusters of lassomycin and lariatin as their cysteine proteases is split into two separate enzymes, which we dubbed, based on their origin, B1 and B2.[28,37,57] The bigger of the two resulting enzymes, B2, is composed of the normally *C*-terminally located protease domain, while the smaller enzyme, B1, features the conserved *N*-terminal domain of the B proteins and shows homology to PqqD, an enzyme that is suggested to mediate the interaction of the PQQ precursor peptide PqqA and its processing enzymes.[3-5,7,23,37,57] A similar function could be conceivable for B1, as previous *in vitro* studies revealed the ATP-dependency of McjB from the MccJ25 biosynthesis system, which is rather unusual for a cysteine protease.[59] Thus, an ATP-consuming chaperone-like function could be proposed for B1 and the *N*-terminal domains of non-split B proteins. Interestingly, a genome mining study also revealed some putative lasso peptide biosynthetic gene clusters in which the encoded B2 protein is directly fused to an ABC-transporter domain, which suggests that in these systems, precursor processing and export of the mature compound could be directly coupled.[23] The aforementioned *in vitro* studies furthermore confirmed the role of McjB as a protease that cleaves off the leader sequence of the McjA precursor peptide and the role of McjC as an ATP-dependent macrocyclase.[59] Surprisingly, these experiments also showed that McjB and McjC are interdependent, as they were proven to be unable to catalyze their corresponding functions on their own, but were capable to perform their catalytic functions when assayed with an inactive variant of the respective other enzyme. Combined, these results led to the postulation of a maturation mechanism for lasso peptides (Figure 1.4). The *in vitro* and an additional *in vivo* study also confirmed residues that are crucial for the activities of McjB and McjC, while another study provided data that suggested McjB and McjC to be at least partially associated with the membrane.[59-61] Nevertheless, the characterization of the MccJ25 processing enzymes was accompanied by extensive problems during their expression and isolation, which is emphasized by the fact that the modification assays required almost stochiometric amounts of the isolated enzymes as well as that the isolated McjB protein was still

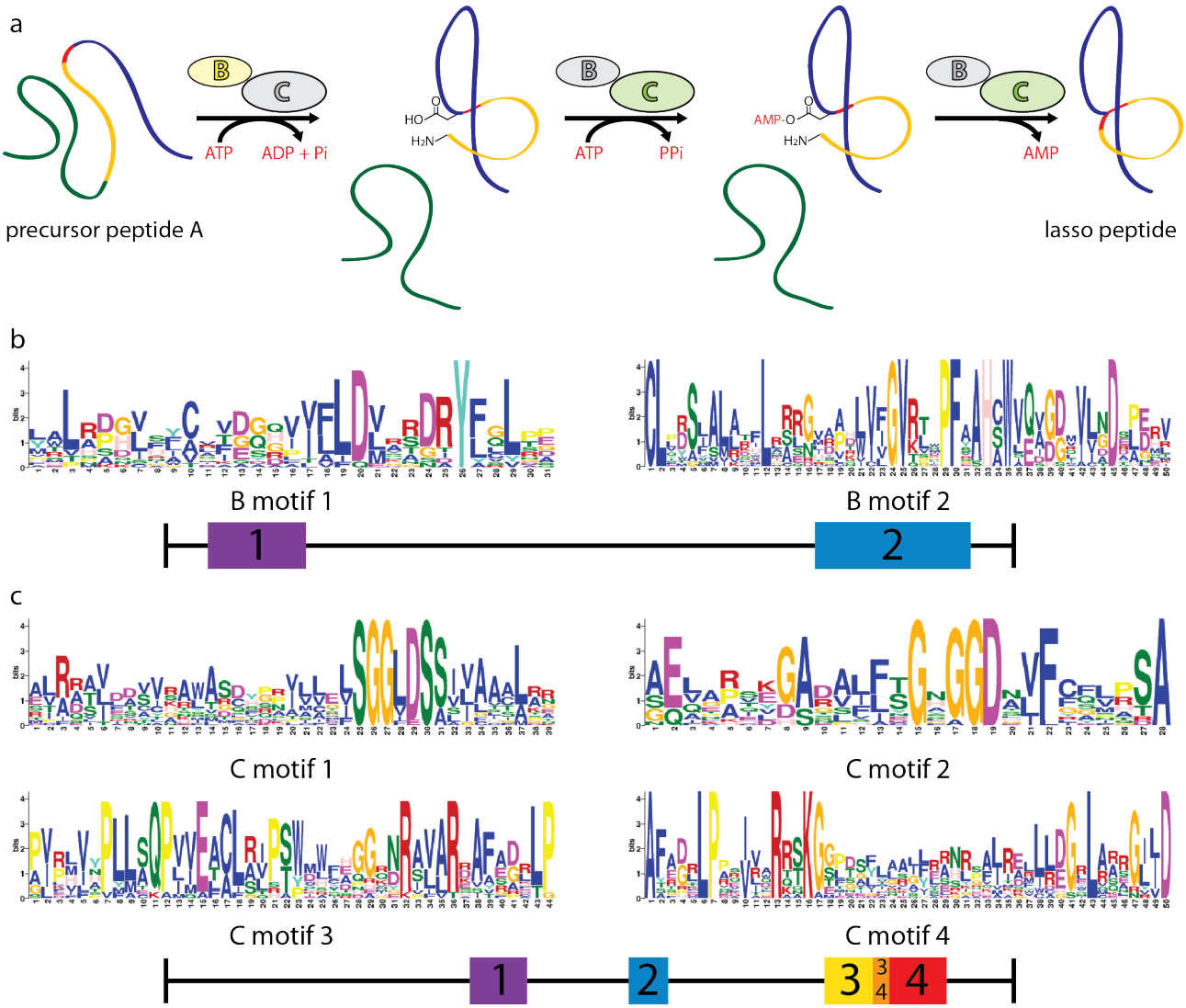


Figure 1.4. Lasso peptide maturation and conserved motifs of the processing enzymes. (a) Schematic depiction of the putative mechanism of lasso peptide biosynthesis. In the first step, the B protein cleaves off the leader peptide under ATP-consumption, which may also mediate the prefolding of the core peptide. In the next step, the C protein activates the carboxyl side chain in an ATP-dependent manner and subsequently catalyzes the macrocyclyzation, yielding the mature lasso peptide.[37,59] (b) Conserved motifs of B proteins.[37] The *N*-terminal domain may be needed for interaction with ATP, while the *C*-terminal domain features a catalytical Cys-His-Asp triad typical for cysteine proteases. In B1 proteins only the N-terminal domain is present, while B2 proteins only contain the cysteine protease domain. (c) Conserved motifs of C proteins.[37] Previous studies[59,61] based on homology comparisons and mutational analysis suggest motifs 1 and 2 to be involved in ATP-binding.

accompanied by significant amounts of impurities.[59] These and similar problems were so far encountered for all maturation enzymes that were tested in our group, suggesting that these enzymes are generally very challenging to produce and isolate. This currently limits the further *in vitro* investigation of the lasso peptide biosynthesis and has so far foiled the possibility to crystallize these enzymes and hence to learn more about their three-dimensional arrangement and how they

actually interact with each other and the precursor peptide. Fortunately, *in vivo* studies involving several lasso peptide biosynthetic gene clusters made it possible to get at least some information about residues of a lasso peptide precursor that are essential for either the recognition by or the activities of the processing enzymes. Besides obvious positions in the lasso peptide sequence itself, like the amino acid at position 1, the ring forming aspartate or glutamate residue and the plug amino acids,[22,25,26,36,47,62] several residues in the *C*-terminal part of the leader sequence have been shown to play a major role for efficient maturation.[22,25,26,36,47,63,64] The most conserved residue in this regard is the threonine at the penultimate position of the leader peptide, which seems so far to be conserved throughout all lasso peptide precursors and shows only restricted tolerance to substitutions, which is then again limited to structurally similar amino acids.[22,25,26,36,47,63,64] A general recognition motif in the *C*-terminal region of the leader peptide, which is at least widely distributed amongst clusters of proteobacterial origin, was also identified very recently[47] and is shown in Figure 1.5. Unlike for the T-2 residue there are also naturally occurring exceptions for this motif known in proteobacterial systems like in the precursor peptides of capistruin and MccJ25. Furthermore, this motif seems to be only loosely conserved in Gram-positive lasso peptide biosynthetic gene clusters.

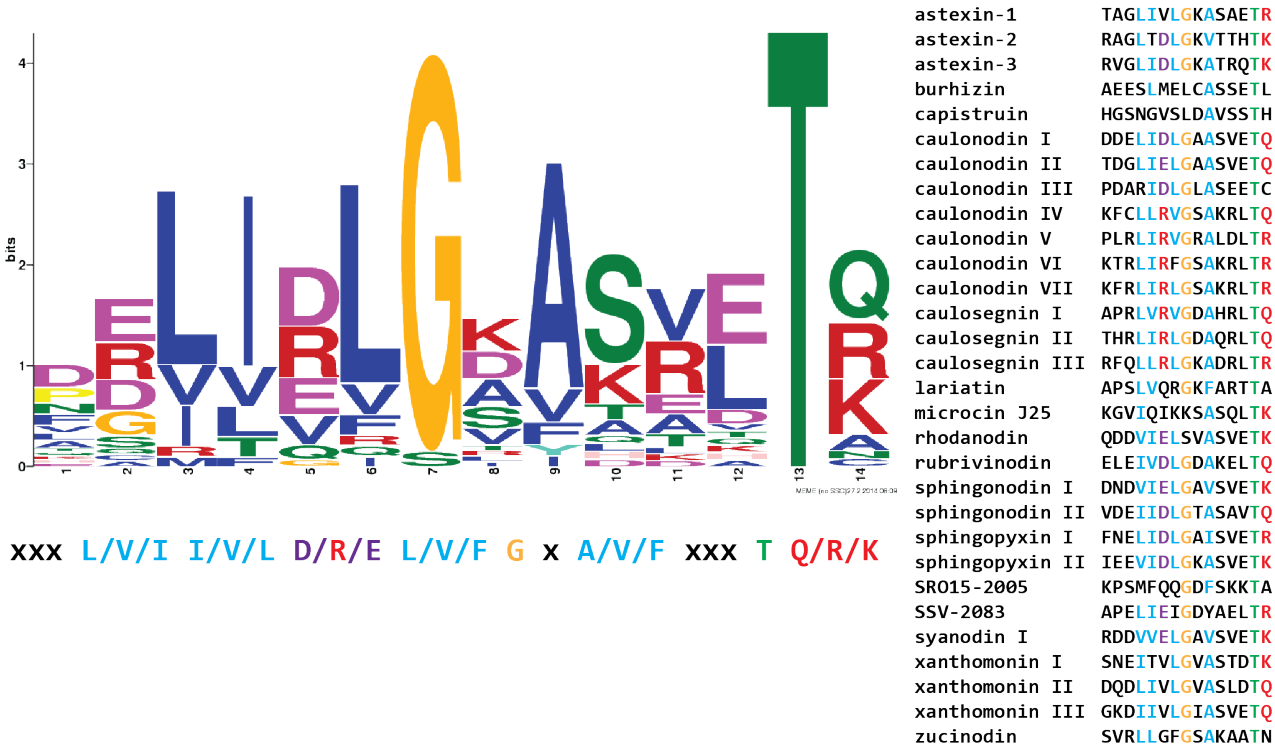


Figure 1.5. On the left, the conserved motif found in the C-terminal regions of the leader peptides of all known functional lasso peptide precursors is shown.[47] The result of the analysis with the MEME algorithm is depicted above, while the derived consensus sequence is shown below. On the right, an alignment of the last 15 amino acids of all leader peptides used for the MEME analysis is shown next to the names of the corresponding lasso peptides. Residues in accordance with the conserved motif are highlighted using the same color code.

It is also interesting to mention that the proteobacterial distribution of precursor peptides that show any alterations in this motif are more or less limited to systems where an ABC-transporter is present.[37,47] This observation may suggest that clusters with D proteins and clusters without D proteins diverged early on in proteobacteria and because of this could be regarded as two subgroups of proteobacterial lasso peptide biosynthetic gene clusters that are only loosely related to each other. With regard to the lasso sequence, several mutational studies were performed *in vivo* with different biosynthetic machineries.[22,25,26,36,47,62,65-69] Taken together, it appears that most of the lasso sequence can be easily exchanged to at least some extent, which speaks for a high promiscuity of the processing enzymes. Still, there are also some residues that show very limited tolerance towards substitution. These are on the one hand some positions that appear to be specific for each individual lasso scaffold and on the other hand some generally crucial residues like the ring forming and the plug amino acids. If such a sensitivity towards substitutions of certain positions arises from problems during the enyzmatic recognition or from factors negatively affecting the stability of the lasso fold can not be satisfyingly concluded by *in vivo* experiments alone, but in a lot of cases the cause for the low or complete lack of production can at least be speculated about based on the nature of the modified residue.

In general, lasso peptides are good targets for genome mining approaches for mainly two reasons. The first is that the ribosomal origin of the precursor peptides allows an easy identification of the amino acid sequences of the mature lasso peptides and hence an accurate prediction of the molecular masses that are to be expected upon lasso peptide production. The second is that the B proteins share little homology to enzymes outside of lasso peptide biosynthetic gene clusters and as such constitute an ideal input for BLAST homology searches to identify so far uncharacterized lasso peptide biosynthetic gene clusters.

1.2.3 Biological Functions of Lasso Peptides

The answer about the function of lasso peptides in nature remains mostly elusive so far. The majority of lasso peptides for which a biological activity is known were identified in culture screens for exactly these functions and therefore most targets identified in this way are of human origin. Even though these activities are interesting from a medicinal perspective, they most likely do not represent the original function of these compounds as these lasso peptides predominantly originated from bacteria living in environments that lack direct contact to humans. An overview over all biological activities known for lasso peptides is shown in Table 1.2.

Table 1.2. Overview over all biological activities known for lasso peptides.

class	lasso peptide	biological activity
I	aborycin, MS-271, NP-06, siamycin I/II, RP71955[10,27,34,40,43]	inhibitor of HIV-replication, inhibitor of calmodulin-activated myosin light chain kinase, antimicrobial activity against *Bacillus subtilis*, *Bacillus brevis*, *Enterococcus faecium*, *Micrococcus luteus*, *Pseudomonas saccharophila*, *Staphylococcus aureus* and *Streptomyces viridochromogenes*
II	anantin[46]	antagonist of atrial natriuretic factor
II	astexin-1[30]	weak antimicrobial activity against *Caulobacter crescentus* CB15
II	capistruin[18]	antimicrobial activity against *Burkholderia* and *Pseudomonas* spp.
II	lariatin[17]	antimicrobial activity against *Mycobacteria* spp. (including *M. tuberculosis*)
II	lassomycin[28]	antimicrobial activity against *Mycobacterium tuberculosis*, *M. avium* subsp. *paratuberculosis* and *M. smegmatis*
II	microcin J25[32]	antimicrobial activity against different *Escherichia*, *Salmonella* and *Shigella* spp.
II	propeptin[41]	inhibitor of prolyl endopeptidase, weak antimicrobial activity against *Pseudomonas aeruginosa*, *Mycobacterium phlei* and *Xanthomonas oryzae*
II	RES-701-1/3[45]	antagonist of endothelin type-B receptor
II	sungsanpin[24]	inhibitor of cell invasion for the human cancer cell line A549
III	BI-32169[48]	antagonist of glucagon receptor

Besides the most likely off target activities, several lasso peptides are known to exhibit antimicrobial activities against certain species of bacteria. Interestingly, these activities are mostly found against strains closely related to or living in the same environment as the producing organisms. It is additionally known that the lasso peptide MccJ25 is not produced before the host organism enters stationary growth phase, at which point nutrients in the culture are becoming scarce.[32,70] Furthermore, the production of MccJ25 was shown to be triggered by ppGpp production, which also correlates with nutrient starvation.[70] Taken together, these observations suggest that such lasso peptides could be produced as a response to onsetting starvation or other stress conditions with the aim to kill off competitors and hence to ensure the survival of the host organism. Interestingly, all known biosynthetic gene clusters that produce lasso peptides with an antimicrobial activity,[18,28,55,57,58] namely the clusters of MccJ25, capistruin, lariatin and lassomycin, also carry a gene encoding a dedicated ABC-transporter, which in case of McjD was directly proven to be essential for conferring immunity to the host organism against MccJ25 by mediating the export of the mature lasso peptide.[55,58] In contrast, the majority of recently reported lasso peptides that were discovered by directed genome mining approaches did not exhibit antimicrobial activities against closely related organisms, but at the same time all of them also lacked an ABC-transporter encoding gene as part of their biosynthetic gene clusters.[22,26,30] This suggests that the presence of an ABC-transporter could be directly linked to the presence of an antimicrobial activity of the produced compound. From all reported lasso peptides with known antimicrobial function, the actual

targets were only found for MccJ25, capistruin and lassomycin.[28,71-77] The in this regard best investigated lasso peptide is MccJ25, whose target was identified by total sequencing of single clones of prior sensitive bacteria that spontaneously developed a resistance against this compound. Thus unsurprisingly, the first uncovered mutations were all located in genes encoding parts of a dedicated import system, namely the enzymes involved in the FhuA-TonB complex that normally mediates the uptake of the iron loaded siderophore ferrichrome.[76,77] Hence, it was proposed that MccJ25 enters the cell in a likewise manner, which was recently confirmed by the elucidation of the co-crystal structure of FhuA and MccJ25 (Figure 1.6) that shows how MccJ25 mimics the binding mode of iron loaded ferrichrome.[78]

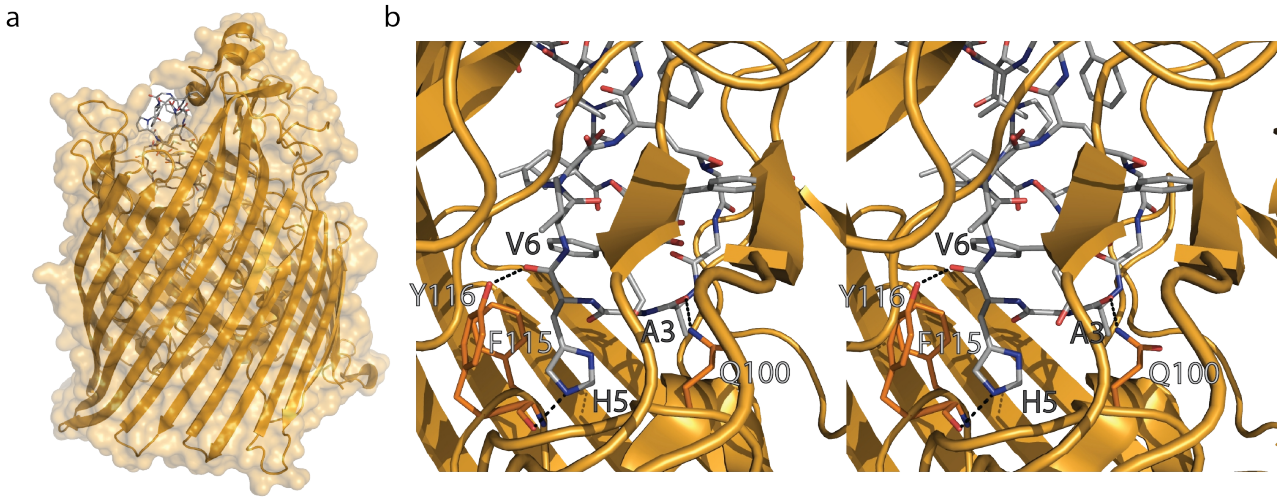


Figure 1.6. Co-crystal structure of the TonB-dependent receptor FhuA with MccJ25 (PDB code 4CU4).[78] (a) Overview of the complete structure. FhuA is depicted in orange as cartoon with overlay of a transparent surface map, while MccJ25 is colored by elements (carbon in gray, nitrogen in blue, oxygen in red). (b) Stereoview of a close-up of the FhuA binding pocket. Ionic interactions are highlighted by black dotted lines. The residues involved in key interactions are shown in light gray for FhuA and in dark gray for MccJ25.

Still, the actual target was only later identified in a more thorough approach and was shown to be the Gram-negative RNA polymerase, which is also true for capistruin.[74,75] Further studies revealed that the antimicrobial activity of MccJ25 is derived from its the ability to inhibit the NTP uptake of the RNA polymerase. The fact that the observed resistance-conferring mutations were all clustered in the secondary channel of the RNA polymerase and the results of subsequent *in silico* modelling experiments both suggest that the lasso peptide is binding to the secondary channel and in this way obstructs the NTP uptake.[72,73] Nonethless, in addition to this activity, a second mode of action for MccJ25 was also found for some highly sensitive *Salmonella* species. In these strains, it was first observed that MccJ25 is apparently able to depolarize the cell membrane and at the same time is

able to both increase cell permeability and decrease oxygen consumption.[79,80] This could later also be shown for *E. coli* strains that overexpressed the FhuA receptor.[81] Interestingly, both modes of action appear to be independent of each other, since it was shown that *E. coli* strains overexpressing FhuA were still susceptible for MccJ25, even when they produced a MccJ25-resistant RNA polymerase or when they were treated with C-terminal amidated MccJ25, which was shown to be unable to inhibit the RNA polymerase *in vitro*.[81,82] Further studies revealed that the antimicrobial activity in these cases is conferred by MccJ25 triggering superoxide production in an yet unknown fashion. This was proven by the fact that strains which both overexpressed FhuA and had a MccJ25-resistant RNA polymerase were sensitive towards MccJ25 under aerobic conditions, but resistant towards MccJ25 under anearobic conditions.[81,83] Furthermore, *in vitro* assays with membrane fractions from *E. coli* showed that in the presence of MccJ25 NADH and lactate dehydrogenase activity was inhibited. This effect could be negated by the addition of a superoxide scavenger molecule, directly linking the observed enzyme inhibition to MccJ25-triggered superoxide formation.[81] In addition, it was shown that for both inhibiting the RNA polymerase and triggering the superoxide formation, MccJ25 had to pass both the periplasmic and the cytoplasmic membranes, as an *E. coli* strain overexpressing FhuA remained resistant towards MccJ25 when carrying a mutation in the SbmA protein, which is proposed to be crucial for the transport of MccJ25 from the periplasm into the cytosol.[77,81,83] Interestingly, a single conserved mutation of Tyr9 to Phe completely abolishes the capability of MccJ25 to cause superoxide production, which strongly suggests Tyr9 to play a crucial role in the underlying mechanism.[83]

The target of lassomycin was detected in a likewise manner to the one of MccJ25. In this case, the resistance-conferring alterations in *Mycobacterium tuberculosis* mutants were clustered in the coding sequence for the *N*-terminal region of ClpC1, an ATPase that interacts with the ClpP1P2 complex and in this way mediates the degradation of proteins by ClpP1P2.[28] Here, the antimicrobial function is achieved by uncoupling of the ATP-hydrolysis activity of ClpC1 from the ClpP1P2-dependent proteolysis, thus not only wasting the cells energy, but at the same time also obstructing the ClpP1P2 activity, which is essential for the viability of *M. tuberculosis*. This shows that even though both MccJ25 and lassomycin have an antimicrobial activity, their mode of action is completely different, highlighting that although all lasso peptides share a common topology, they may indeed be quite diverse in regard to their biological functions. The co-occurrence of antimicrobially active lasso peptides with D protein encoding genes in their biosynthetic gene clusters furthermore suggests that lasso peptides lacking such an ABC-transporter system may instead have a completely different function, which could be revealed by closer inspection of the surroundings of corresponding gene clusters.

1.2.4 Lasso Peptide Isopeptidases and TonB-dependent Receptors

Closer inspection of the surroundings of the ABC-type caulosegnin biosynthetic gene cluster revealed that downstream to it another reverse facing gene cluster was present that contained four genes. These genes putatively encode a σ-factor, an anti-σ-factor, a TonB-dependent receptor and a peptidase.[22,37] Analogous gene clusters were also found adjacent to both lasso peptide biosynthetic gene clusters from *Asticaccaulis excentricus*.[22,23,25,37] Based on these observations, Link and co-workers inspected the predicted peptidases from *A. excentricus* more closely and showed that these enzymes were indeed dedicated lasso peptide isopeptidases.[23] *In vitro* studies with the isopeptidase close to the second astexin biosynthetic gene cluster revealed that this enzyme cleaved the isopeptide bonds in the macrolactam rings of astexin-2 and -3 with high specificity, while a branched-cyclic peptide with the same primary structure as astexin-2 was completely resistant towards this enzyme. Interestingly, the isopeptidase showed no activity against astexin-1. Based on a bioinformatic sequence analysis, a catalytic triad consisting of Ser527-Glu610-His638 was identified and it was proven that a S527A mutation disrupts the catalytic activity of this enyzme, providing a strong evidence that these isopeptidases indeed belong to the family of serine proteases. Additionally, *in vivo* studies could confirm that astexin-1 was selectively degraded by the isopeptidase adjacent to its own biosynthetic gene cluster. Taken together, these experiments show that these isopeptidases recognize and degrade only the lasso peptides that are produced by the corresponding adjacent ABC-type biosynthetic gene clusters with a very high specificity, suggesting a general involvement of the other enzymes from the isopeptidase gene cluster with lasso peptides as well.[23] This is especially intriguing considering that FhuA, which was shown to mediate the MccJ25 uptake in strains sensitive towards this lasso peptide, also belongs to the family of TonB-dependent receptors.[76-78] Interestingly, the similarities between lasso peptide and FhuA-like systems do not end here, as the σ- and anti-σ-factors show homology to FecI and FecR, respectively.[22,23,37] FecI and FecR are regulatory enzymes associated with another siderophore uptake system similar to the one FhuA belongs to (Figure 1.7).[84-86]

Hence, it is reasonable to assume that the enzymes encoded by the isopeptidase gene clusters could have similar functions involving the respective lasso peptides. While the σ- and anti-σ-factors could regulate the transcription of certain promoters in response to the presence or absence of the corresponding lasso peptide, the TonB-dependent receptor could mediate the uptake of this lasso peptide into the cell, where it then would be linearized by the isopeptidase. Based on this, two possible functions for lasso peptides from a proteobacterial ABC-type biosynthetic gene cluster seem plausible. Either, they could have a scavenging function likewise to siderophores and would

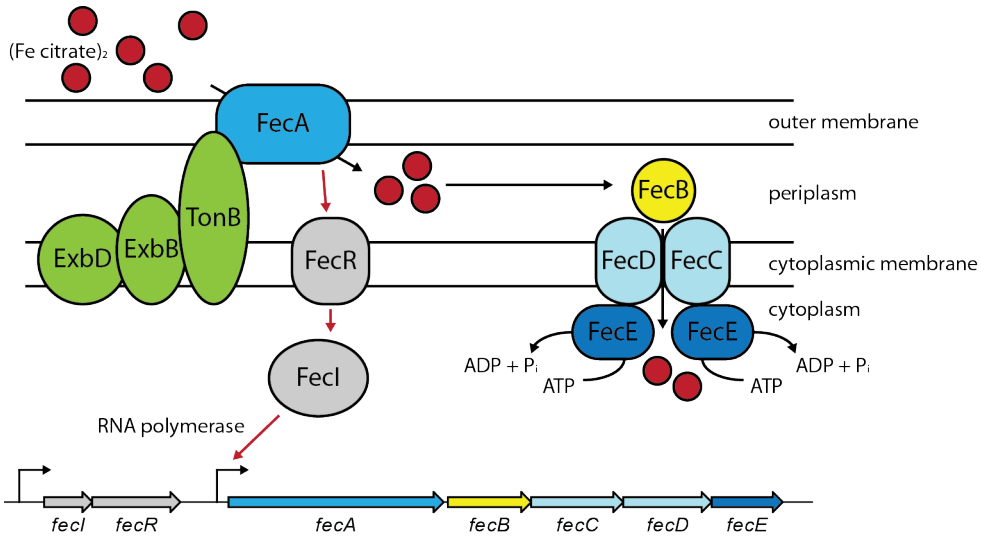


Figure 1.7. Schematic representation of the ferric citrate uptake in *E. coli* K12 mediated by the Fec proteins.[84-86] First, extracellular ferric citrate (red spheres) binds to the TonB-dependent receptor FecA. FecA then mediates both the uptake of ferric citrate into the periplasm by interaction with the TonB protein complex (black arrow) as well as the activation of FecR (red arrow). Subsequently, activated FecR binds FecI and thus allows interaction of the FecI σ-factor with the RNA polymerase, which initiates a strong transcription of the *fecABCDE* operon. In the periplasm, ferric citrate can interact with a protein complex consisting of the perisplasmic binding protein FecB, the transmembrane proteins FecC and FecD and the cytoplasmic ATPase FecE. The complex then facilitates the transport of ferric citrate through the cytoplasmic membrane under ATP-consumption.

release their scavenged molecules or ions upon degradation by the isopeptidases, or they could play a role as signaling molecules. From these proposed functions, the scavenging role is rather unlikely as all lasso peptides may share their basic three-dimensional appearance, but have a highly diverse primary structure showing no conservation of any particular residues that could be involved in selective binding of a target molecule or ion. Therefore, a signaling function seems more feasible. In this scenario, the linearization of the lasso peptide by the isopeptidase would either function as a way to slowly diminish the signal in case no new lasso peptide is internalized or, on the contrary, the signaling function would not be conferred by the lasso peptide itself, but only by the resulting linear peptide. Of course, all of this is purely speculative at the moment and needs further testing to conclusively reveal the real function of these lasso peptides. Nevertheless, these or similar functions would be in agreement with the observed lack of antimicrobial activities of the lasso peptides from proteobacterial ABC-type biosynthetic gene clusters, which points to a different role in nature as merely killing off competing organisms under nutrient limiting conditions. The fact that two separate genome mining studies could furthermore show that most of the ABC-type clusters from proteobacteria have a neighboring isopeptidase gene cluster, additionally points out that the corresponding lasso peptides probably all fulfill a similar function in their native hosts.[23,37]

1.3 Epitope Grafting

1.3.1 Advantages and Drawbacks of Peptides in Medicinal Applications

In nature, there are many instances known where peptides act as signaling molecules, e.g. hormones like somatostatin, somatropin or insulin. Targeting the corresponding receptors, of which some use only small peptide epitopes for substrate recognition, can induce or inhibit certain cell functions, which can be of special medicinal interest for affecting pathogens, but also for affecting cancerous and non-cancerous human cell lines. Additionally, peptides became of interest for potential therapeutic applications in the context of other targets, e.g. as inhibitors of specific enzymes or for disturbing certain protein-protein interactions and thus the functions of the corresponding multienzyme complexes.[87-111] For discovery of peptides with such activities, methods for high-throughput generation and screening of vast compound libraries have been developed and successfully employed for the identification of peptide based drug-leads.[8,96-98,101,102,104,106,112] Nonetheless, the major drawback of linear peptides is that they are highly susceptible to proteolytic degradation, which results in a very low half-life in the human body, making them unsuitable for therapeutic applications.[65,87,88,91,105,108,110,113] A solution for this problem is to incorporate the peptide pharmacophore into more stable scaffolds. These can be as simple as small, constrained cyclic peptides, whose stability against proteolytic degradation can be further improved by the incorporation of D-amino acids. Prominent examples of this are cilengitide and octreotide.[87-89,91,94,100,104,106,109,112-116] The former compound is an anti-cancer agent currently under investigation in clinical trials that selectively binds to the αvβ3 integrin receptor and through this is able to inhibit angiogenesis, which is a crucial factor for tumor growth and metastasis. The latter compound is a somatostatin receptor antagonist that is approved as a drug for either tumor therapy or tumor imaging, depending on which chelator moiety and corresponding complexed radionuclide is linked to the peptide scaffold. Nevertheless, there are more and more examples known, were complex and highly stable natural peptide scaffolds were successfully employed for presenting a bioactive peptide epitope to an interaction partner.

1.3.2 Cyclotides as Potent Scaffolds for Epitope Grafting Approaches

The most prominent examples for natural scaffolds used in epitope grafting efforts are probably the cyclotides.[93,95,105,108,110,117-124] These natural products are plant-derived, consist of around 30 amino acids and feature the unique structural motif of the so-called cyclic cystine knot (CCK), which consists of macrocyclized *C*- and *N*-termini combined with three disulfide bonds that form a knotted and very compact structure. This structural motif has a very promiscuous primary structure and at

the same time confers a high stability against proteolytic and thermal degradation, therefore making cyclotides ideal candidates for the delivery and presentation of peptide epitopes to protein targets *in vivo*. The general appearance of such a structure is shown exemplary for the first cyclotide discovered, namely kalata B1,[117] in Figure 1.8.

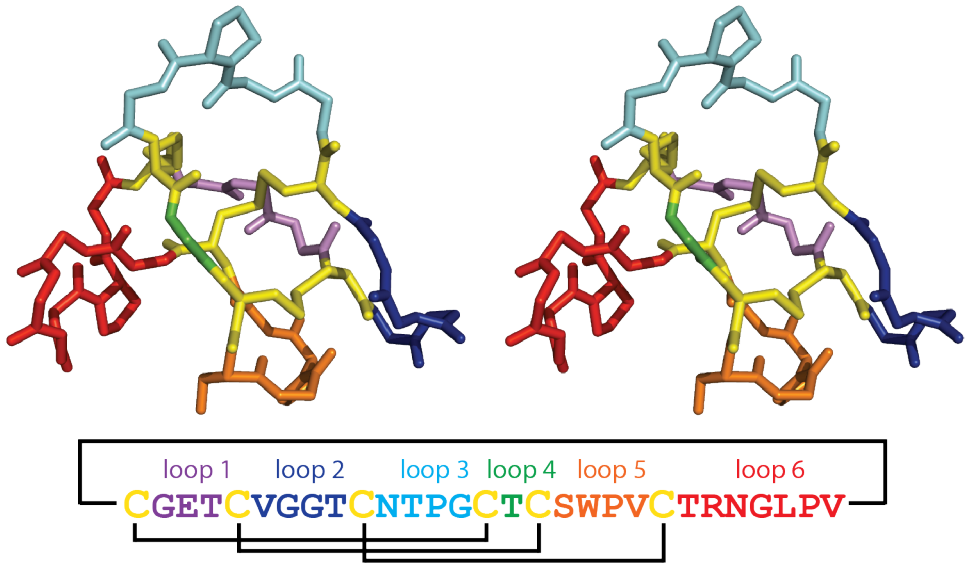


Figure 1.8. Stereoview of the three-dimensional structure of the cyclotide kalata B1 (PDB code 1KAL).[117] Depicted below is the primary structure of this natural product using the same color scheme for the corresponding residues as above.

State of the art research employs solid phase synthesis, native chemical ligation and heterologous expression utilizing intein chemistry for production of novel cyclotide variants. These methods not only provide efficient ways to access a large cyclotide diversity, but also generally allow the incorporation of unnatural amino acids. So far, cyclotide grafts were generated that were able to act as inhibitors for proteases like β-tryptase and the human leukocyte elastase,[119,120] as agonists or antagonists for the vascular endothelial growth factor,[118] melanocortin 4,[121] bradykinin B1[123] and several integrin receptors[105,122] or to negate the binding of Hdm2 and HdmX to p53 and thus reactivating the p53 tumor suppressor pathway in certain cancer cell lines.[108] These results in combination with the easy access to cyclotide scaffolds by synthetic approaches as well as biosynthesis highlights the potential of cyclotides in epitope grafting endeavors and explains why the cyclotide scaffold is so prominent in this research area.

1.3.3 Lasso Peptides as Potent Scaffolds for Epitope Grafting Approaches

Similar to cyclotides, the extraordinary stability of the lasso fold[9,10,15,17,22,25-27,31-39] predisposes these compounds for use in epitope grafting efforts. In contrast to cyclotides, up to now the potential of lasso peptides as scaffolds in this field of research was only barely investigated.[65] From all lasso peptides discovered so far, MccJ25 is probably the most suitable scaffold for epitope grafting approaches (Figure 1.9).

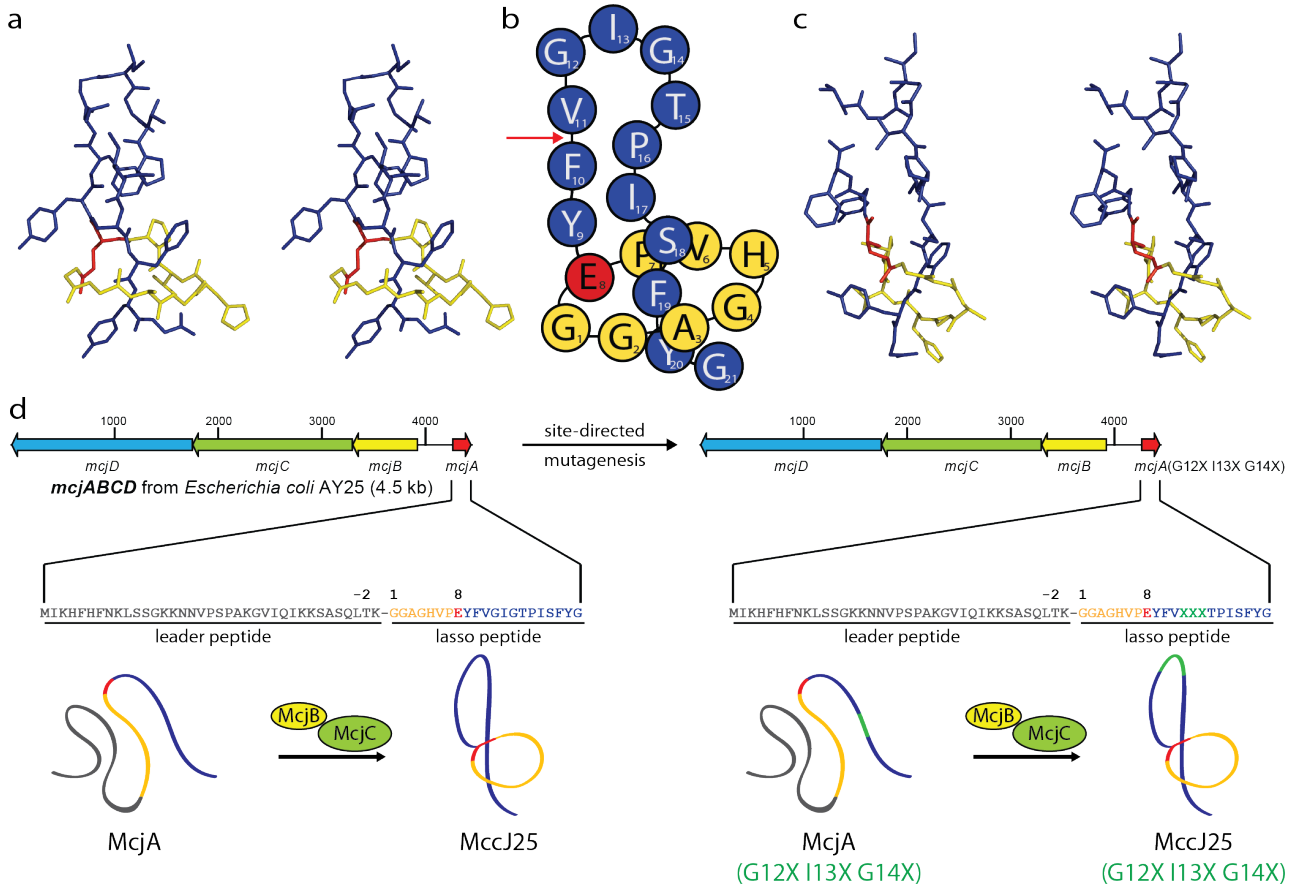


Figure 1.9. MccJ25 as scaffold for epitope grafting approaches. The ring forming glutamate residue is colored in red, the remaining ring amino acids in yellow and the tail amino acids in blue. (a) Stereoview of the three-dimensional structure of MccJ25 (PDB code 1Q71).[14-16] (b) Schematic representation of MccJ25. The red arrow indicates the thermolysin cleavage site. (c) Stereoview of the three-dimensional structure of thermolysin cleaved MccJ25 (PDB code 1S7P).[49] (d) Schematic representation of the grafting of the fictional XXX peptide epitope into the loop of MccJ25 replacing Gly12, Ile13 and Gly14. The incorporated mutation is highlighted in green.

One reason for this is the extraordinary high stability of MccJ25, as it was reported that this lasso peptide retains its antimicrobial activity even after autoclaving.[32] Furthermore, MccJ25 is the only lasso peptide known that can be converted into a stable two-chain peptide complex by either thermolysin treatment (Figure 1c)[35,39,49,50] or gas phase fragmentation.[15,16,50,67] In this state, the tail and

the ring of the lasso peptide are no longer covalently linked, but are still interlocked so tightly that they form a single molecular unit that retains its ability to inhibit the Gram-negative RNA polymerase.[39] The fact that the biosynthetic gene cluster of MccJ25 originated from an *E. coli* strain additionally facilitates the heterologous production of this lasso peptide and variants thereof in standard *E. coli* lab strains, which provides an efficient way to access modified MccJ25 scaffolds by biosynthesis. This is especially important as until now lasso peptides remain inaccessible by means of chemical synthesis. Another important criterion that qualifies MccJ25 as promising scaffold for epitope grafting efforts is its large and exposed loop region, which is important for the incorporation and presentation of a peptide pharmacophore. With eleven tail amino acids above the macrolactam ring, this is by far the largest loop of all lasso peptides with so far known three-dimensional structures.Compared to cyclotides, lasso peptides may appear inferior as scaffolds for epitope grafting endeavors as their limitation to biosynthetic production complicates the incorporation of unnatural amino acids as well as the production of large compound libraries, which necessitates a more rational approach to the epitope grafting with these kind of scaffolds. Besides these limitations, lasso peptides also have advantages over cyclotides for medicinal applications. Namely, they are smaller and have a more compact structure and since their fold does not rely on disulfide bonds, they are in contrast to cyclotides resistant towards reductive conditions.

Before this thesis, only one example of epitope grafting employing a lasso peptide scaffold was reported.[65] In this study, the integrin recognition sequence RGD was incorporated into the loop of MccJ25. This lasso graft was shown to be able to bind the αvβ3 integrin *in vitro* as well as being able to inhibit the capillary formation of human umbilical vein endothelial cells (HUVECs) in cell culture assays. Still, even though the compound had a promising affinity in the nanomolar range for binding to αvβ3 *in vitro*, the selectivity for this integrin receptor was rather poor, making the lasso peptide unsuitable for therapeutic use. Nonetheless, this study served as a proof-of-concept that established lasso peptides as promising epitope grafting scaffolds and the reported lasso graft can serve as a starting point for the development of a more potent and more selective lasso peptide based integrin inhibitor.

2. Motivation and Objectives

Lasso peptides are a family of RiPPs that were first discovered in the beginning of the 1990s. For more than a decade after their discovery, the isolation of new representatives of this natural product family was decided by chance and because of this, little was initially known about their biosynthesis, characteristics and functions. This changed with the emergence of increasingly inexpensive sequencing techniques through which a vast amount of genomic data became accessible to the scientific community. With this steadily growing wealth of information, bioinformatic analysis of genetic data allowed an efficient way to assess the distribution and prevalence of certain genes and their corresponding proteins in nature. Furthermore, it allowed the genome mining for natural product biosynthetic gene clusters in the sum of the available genomic data and in this way facilitated the targeted search for systems producing novel representatives of natural product families. Amongst many others, this approach can be employed for lasso peptides and thus in 2008 the first lasso peptide isolated in the course of a genome mining study was reported.[18] After this initial proof-of-principle, the next logical step was to optimize the overall procedure to allow a fast and efficient way for the isolation and subsequent characterization of novel lasso peptides. This was the primary objective of this thesis and was targeted at not only increasing the overall number of known representatives of this natural product family, but at the same time these studies should also yield new information about their biosynthesis, physcio-chemical properties and their role in nature.

The second objective of this thesis was to show that the extraordinary stability of these compounds could be combined with specific peptide epitopes that confer medicinal relevant biological activities. In this regard, the aim was to use a previously reported lasso graft that acted as inhibitor of the $\alpha v\beta 3$ integrin receptor[65] and to improve its bioactivity by rational means.

3. Results

This section contains an overview over all studies published in the course of this thesis. This overview includes a description of the author contributions and a short background and summary for each study alongside the publications themselves. The corresponding Supporting Information can be found in the appendix.

3.1 Caulosegnins I-III: A Highly Diverse Group of Lasso Peptides Derived From a Single Biosynthetic Gene Cluster

Julian D. Hegemann, Marcel Zimmermann, Xiulan Xie, and Mohamed A. Marahiel, Caulosegnins I-III: A Highly Diverse Group of Lasso Peptides Derived From a Single Biosynthetic Gene Cluster, *J. Am. Chem. Soc.* **2013**, *135(1)*, 210-222.
doi: 10.1021/ja308173b

Author contributions:

The project was designed by Julian D. Hegemann and Mohamed A. Marahiel with the help of Marcel Zimmermann. All experiments were carried out by Julian D. Hegemann with the exception of the NMR structure elucidation of caulosegnin I, which was done by Xiulan Xie. Analysis and interpretation of the experimental data was done by Julian D. Hegemann with assistance provided when needed by Marcel Zimmermann. The manuscript was written by Julian D. Hegemann and revised with the support of Mohamed A. Marahiel and Marcel Zimmermann. Mohamed A. Marahiel was furthermore responsible for the project supervision in general.

Background and Summary:

Several lasso peptides were known prior to this work, but information about the genetic background of their biosynthetic machineries were scarce. At this point, only four functional lasso peptide biosynthetic gene clusters were known, namely the ones producing MccJ25,[55,56,58] capistruin,[18] lariatin[57] and astexin-1.[30] The major problems preceding this publications persisted since the isolation and characterization of capistruin by Marahiel and co-workers.[18] Capistruin was the first lasso peptide that was isolated using a genome mining approach that was directed at finding and isolating a novel lasso peptide, while lasso peptides that were isolated before, where just found in screenings for compounds with specific bioactivities. What makes capistruin special is that its homologous production by cultivating *Burkholderia thailendenis* E264 in M20 minimal medium was already sufficient for allowing the isolation of this lasso peptide in amounts suitable for characterization and NMR structure elucidation. At the same time, the heterologous production of this lasso peptide in *E. coli* only provided trace amounts, which could only be used for basic MS analysis. Since then, no other lasso peptide investigated in this group showed a homologous production anywhere close to capistruin, which necessitated the optimization of the heterologous lasso peptide production in *E. coli*. In this work, results from Link and co-workers[64,125] were used as a starting point to develop a robust and efficient way for generally optimizing the yields of heterologous lasso peptide production by expression of proteobacterial gene clusters in *E. coli*. This not only allowed the isolation and characterization of caulosegnins I-III in this study, but at the same time this was the basis that made all the subsequent genome mining studies of lasso peptides that were performed in our group possible.[26,37,47] Additionally, this work showed for first time that a single lasso peptide biosynthetic gene cluster could contain more than one precursor encoding gene and that it had the capability to produce a quite diverse set of lasso peptides, which even differed in the sizes of their macrolactam rings. A thorough mutational analysis furthermore provided new information about the lasso peptide biosynthesis and surprisingly revealed that smaller amino acids as previously assumed are suitable to maintain the lasso fold of caulosegnin I. In addition to these results, this study also disproved the former notion that, as it is true for MccJ25,[32] all lasso peptide would have an instrinsic thermal stability, which was made even more intriguing by the fact that caulosegnins I and III were shown to be heat labile, while caulosegnin II was proven to be a heat stable lasso peptide. Thus, this study was the first to come that made way to revising the former paradigms of lasso peptide research.

JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY
Article
pubs.acs.org/JACS

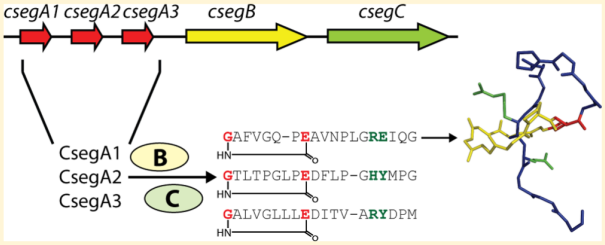
Caulosegnins I–III: A Highly Diverse Group of Lasso Peptides Derived from a Single Biosynthetic Gene Cluster

Julian D. Hegemann, Marcel Zimmermann, Xiulan Xie, and Mohamed A. Marahiel*

Department of Chemistry, Biochemistry, Philipps-University Marburg, Hans-Meerwein-Strasse 4 and LOEWE-Centre for Synthetic Microbiology, D-35032, Marburg, Germany

Supporting Information

ABSTRACT: Lasso peptides are natural products of ribosomal origin with a unique knotted structural fold. Even though only a few of them are known, recent reports of newly isolated lasso peptides were scarce. In this work, we report the identification of a novel lasso peptide gene cluster from *Caulobacter segnis*, that produces three new lasso peptides (caulosegnins I, II, and III) using a single biosynthetic machinery. These lasso peptides possess different ring sizes and amino acid sequences. In this study, we have developed a system for enhanced lasso peptide production to allow isolation of these compounds through heterologous expression in *Escherichia coli*. We were able to elucidate the structure of the most abundant lasso peptide caulosegnin I via NMR spectroscopic analysis and performed a thorough mutational analysis that gave insight into their biosynthesis and revealed important factors affecting the stabilization of the lasso fold in general. The caulosegnins also show a diverse behavior when subjected to thermal denaturation, which is exceptional as all lasso peptides were believed to have an intrinsic high thermal stability.



INTRODUCTION

Lasso peptides are ribosomally assembled post-translationally modified peptides that show an inimitable knotted fold. They feature a macrolactam ring formed between the N-terminus of a glycine or cysteine residue at position 1 and the side chain of a glutamate or aspartate residue at position 8 or 9.[1−13] Additionally, they have a C-terminal tail that is threaded through the macrolactam ring and sterically trapped by bulky amino acids.[3,6,11,12,14−20] This structural motif is the main reason why lasso peptides often exhibit an extraordinary stability against heat, chemicals and proteolytic degradation[2,3,6,8,11,21−25] and it is worth mentioning that up to date it is still not possible to generate such structures through synthetic means.[26,27] Known lasso peptides can be divided into three classes. Class I lasso peptides feature a cysteine residue at position 1 and have a total of four cysteine residues forming two disulfide bonds. Class II lasso peptides feature a glycine residue at position 1 and do not contain any cysteines. Class III lasso peptides combine features of the first two classes, containing a glycine at position 1 and two cysteines forming one disulfide bond.[14]

In addition to their unique structural fold, lasso peptides often possess interesting biological activities such as inhibition of the HIV-protease,[6,7] the glucagon receptor,[10] the atrial natriuretic factor[1] and other medicinally relevant targets,[4,5,8,9] as well as antimicrobial activities.[2,3,7−9,11,12,28−31] While the antimicrobial activities, which are often directed against closely related species or bacteria living in the same habitat, are assumed to give the producer an advantage against competitors when nutrients are scarce, the anti-enzymatic activities may not reflect the natural functions of the lasso peptides, as they were observed in systematic screens for compounds with such activities.

Lasso peptides are interesting not only for their intrinsic bioactivities, but also for their genetic accessibility and their highly promiscuous processing enzymes, allowing their use for epitope grafting. It was recently shown that microcin J25 (MccJ25), the best studied lasso peptide, could be utilized as a scaffold to carry the RGD peptide epitope, enabling a nanomolar inhibition of the $\alpha_v\beta_3$ integrin receptor that plays major roles in angiogenesis and tumor growth.[32]

Even though lasso peptides feature such interesting characteristics and activities, the biosynthetic gene clusters of only four lasso peptides (as well as two putative clusters from lasso peptides only observed via mass spectrometric analysis) are currently known.[12,13,33−35]

The gene cluster of MccJ25 contains four genes. The *mcjA* gene encodes the precursor peptide, while *mcjB* encodes an ATP-dependent cysteine protease with homology to transglutaminases, which was shown to cleave off the leader peptide and is assumed to have some chaperone activity because of its ATP-dependency.[36−38] An asparagine synthetase homologue is encoded by *mcjC*, which catalyzes the ring formation in an ATP-dependent manner,[36−38] whereas *mcjD* encodes an ABC-transporter, that confers immunity to its host.[33] The capistruin

Received: August 22, 2012
Published: December 8, 2012

biosynthetic gene cluster is arranged in a similar manner,[12] while that of lariatin contains an additional gene coding for an 84 amino acid small protein of unknown function that was shown to be essential for lasso peptide maturation.[34] The gene cluster of the recently reported astexin-1 contains homologues to *mcjB* and *mcjC*, but unlike all clusters mentioned above lacks an ABC-transporter.[35] As the ATP-dependent cysteine proteases have no known homologues outside of lasso peptide biosynthetic gene clusters, they are useful targets for genome mining approaches.[12]

Through use of a genome mining strategy combined with an optimized heterologous expression system, we were able to identify a new lasso peptide biosynthetic gene cluster from *Caulobacter segnis* and could show that this single enzymatic system is able to produce three class II lasso peptides, designated caulosegnins I–III, with different primary structures, ring sizes and stability profiles. Furthermore, we provide the NMR structure of caulosegnin I and a detailed mutagenesis study of this novel system.

EXPERIMENTAL PROCEDURES

Bacterial Strains and General Methods. *C. segnis* (DSM no. 7131) was purchased from the German Collection of Microorganisms and Cell Cultures (DSMZ). *Escherichia coli* TOP10 was used for cloning and *E. coli* BL21 (DE3) was used for heterologous expression. Both strains were purchased from Invitrogen. DNA dideoxy sequencing confirmed the identity of constructed plasmids and mutants thereof. Oligonucleotides, proteinase K and carboxypeptidase Y were purchased from Sigma Aldrich. Elastase, trypsin and chymotrypsin were purchased from Promega. Restriction enzymes, Phusion polymerase and T4 DNA ligase were purchased from New England Biolabs.

Fermentation of *C. segnis*. For homologous fermentation of *C. segnis*, different culture media were used, namely, LB medium (10 g/L bactotryptone, 5 g/L yeast extract, 5 g/L NaCl, pH = 7.0), M9 minimal medium (17.1 g/L $Na_2HPO_4 \cdot 12\ H_2O$, 3 g/L KH_2PO_4, 0.5 g/L NaCl, 1 g/L NH_4Cl, 1 mL/L $MgSO_4$ solution (2 M), 0.2 mL/L $CaCl_2$ solution (0.5 M), pH = 7.0, after autoclaving, 10 mL/L glucose solution (40% w/v) and 0.2 mL/L vitamin mix (Supporting Information Table S1) were added), M20 minimal medium (20 g/L L-glutamic acid, 0.2 g/L L-alanine, 1.0 g/L sodium citrate, 20 g/L $Na_2HPO_4 \cdot 12\ H_2O$, 0.6 g/L KCl, 0.6 g/L Na_2SO_4, 0.02 g/L $FeSO_4 \cdot 7\ H_2O$, 2 mL/L $MgCl_2$ solution (0.5 M), 103 μL/L $CaCl_2$ solution (0.5 M), 346 μL/L $MnSO_4$ solution (0.1 M), pH = 7.0, after autoclaving, 0.2 mL/L biotin solution (10 mg/L) and 0.2 mL/L thiamine solution (10 mg/mL) were added), M63 minimal medium (13.6 g/L KH_2PO_4, 2 g/L $(NH_4)_2SO_4$, 1 g/L bactopeptone, 1 mL/L $MgSO_4$ solution (2 M), pH = 7.0, after autoclaving, 5 mL/L glucose solution (40% w/v), 0.2 mL/L biotin solution (10 mg/L) and 0.2 mL/L thiamine solution (10 mg/mL) were added) and PYEGR medium (2 g/L bactopeptone, 1 g/L yeast extract, 0.8 mL/L $MgSO_4$ solution (2 M), pH = 7.0, after autoclaving, 2.5 mL/L glucose solution (40% w/v) and 0.2 mL/L riboflavin (2 mg/mL) were added).

A total of 300 mL of each medium was inoculated with 3 mL of *C. segnis* PYEGR overnight cultures and cultivated for 7 days at 20 or 30 °C. Cultures were harvested by centrifugation. Cell pellets were extracted with 50 mL of MeOH. Culture supernatants were applied to solid phase extractions using XAD16 resin (Sigma Aldrich, ~7 g/L culture supernatant). After 1 h of stirring the XAD16 containing supernatant at room temperature, the supernatant was removed by filtration. The resin was then washed with water and subsequently eluted with a total of 50 mL of MeOH. Solvent was removed via evaporation at 40 °C and reduced pressure. Dry pellet and supernatant extracts each were resuspended in a total of 900 μL of 50% MeOH, cleared by centrifugation and analyzed via LCMS.

Mass Spectrometric Analysis. Mass spectrometric analysis of extracts was performed with a LTQ-FT ultra instrument (Thermo Fisher Scientific) connected to a microbore 1100 HPLC system (Agilent) using 100 μL extract for each measurement. For detection, the UV absorption at 215 nm was recorded. Separation was achieved using a CC 125/2 Nucleosil 300–8 C18 column (Macherey-Nagel) applying the following gradient of water/0.1% trifluoroacetic acid (solvent A) and MeCN/0.1% trifluoroacetic acid (solvent B) at a column temperature of 40 °C and a flow rate of 0.2 mL/min: holding 2% B for 2 min, followed by a linear increase from 2% to 30% B in 18 min, a subsequent linear increase from 30% to 95% B in 15 min and holding 95% B for additional 2 min.

Mass spectrometric analysis of the thermal stability and protease assays of the purified lasso peptides was performed with a 1100 series MSD (Hewlett-Packard) coupled with a microbore 1100 HPLC system (Agilent). Separation was achieved using a EC 125/2 Nucleodur 100-3 C18ec column and specific gradients. For caulosegnin I the following gradient was used at 50 °C column temperature at a flow rate of 0.2 mL/min: Linear increase from 25% to 35% B in 15 min, followed by a linear increase from 35% to 95% B in 2 min and holding 95% B for additional 5 min. For caulosegnin II and III, another gradient at 25 °C column temperature and a flow rate of 0.3 mL/min was used: Linear increase from 30% to 50% B in 15 min, followed by a linear increase from 35% to 95% B in 2 min and holding 95% B for additional 5 min.

Collision-induced dissociation fragmentation studies within the linear ion trap were done using online LCMS. In all cases, the doubly charged ions were selected for fragmentation as they were the dominant species in the spectrum. The energy for fragmentation was set to 35% for every measurement performed.

For quantification UV-peak areas were integrated and relative production was determined by comparison to the UV-peak area of the corresponding wild type.

Cloning, Mutagenesis, and Modification of the Caulosegnin Gene Cluster. For cloning, the *csegAAABC* gene cluster was amplified from genomic DNA of *C. segnis* by PCR using appropriate oligonucleotide primers (Supporting Information Table S2), Phusion polymerase, GC-buffer and 10% DMSO. Annealing temperature was set to 65 °C and extension was performed for 4.5 min at 72 °C. After purification, the resulting ~3.2 kb amplicon was digested with *NdeI* and *HindIII* and cloned into the likewise digested pET41a vector.

Mutagenesis was performed via PCR using the above-mentioned conditions and 5′-phosphorylated oligonucleotide primers (Supporting Information Table S3–S5) containing the mutated sequence. The linear DNA amplicon (carrying the desired mutation) was purified, subjected to *DpnI* digestion to remove residual template DNA and a blunt end ligation at 16 °C overnight was performed. For ligation, 150–500 ng of purified DNA amplicon was used. Deletions and sequence substitutions were achieved in a likewise manner with appropriate 5′-phosphorylated oligonucleotide primers (Supporting Information Table S6).

Ligated DNA was transformed into *E. coli* TOP10. After overnight incubation on LB agar plates containing 50 μg/mL kanamycin, plasmids were prepared from indiviual transformants and isolated plasmids were analyzed by DNA sequencing. For heterologous expression, plasmids bearing the correct sequences were retransformed into *E. coli* BL21(DE3).

Heterologous Synthesis of Caulosegnins I–III in *E. coli*. LB medium was supplemented with 50 μg/mL kanamycin, inoculated with *E. coli* BL21 cells carrying the respective plasmids and grown overnight at 37 °C. With such overnight cultures, the fermentation media were inoculated to an OD_{600} of 0.01.

Initial screens were done in M9, M20 and M63 minimal medium and in LB medium and fermented for either 1 day at 37 °C or 3 days at 20 °C. Expression was induced by the addition of IPTG to a final concentration of 0.05 mM at an OD_{600} of ~0.6. For fermentation at 20 °C, cells were at first grown to an OD_{600} of ~0.4 at 37 °C, then slowly cooled to 20 °C in the course of 1 h, induced when reaching the target OD_{600} of ~0.6 and furthermore cultivated at 20 °C until harvested.

For optimization screenings, 1200 mL of M9 minimal medium were used for each construct. For variants (and respective wild type controls) 600 mL M9 minimal medium were used for each variant. In

both cases, the cultures were fermented as described for 3 days at 20 °C. At the end of each fermentation of a variant, the OD_{600} was measured and compared to the OD_{600} of the corresponding wild type fermentation to rule out that a possible toxicity of the produced compound hindered the cell growth and thus caused a low yield. Cells were harvested by centrifugation and cell pellets were extracted with 50 mL of MeOH. Solvent was removed at 40 °C and reduced pressure. Dried extracts were resuspended in a total of 800 μL of 50% MeOH, cleared by centrifugation and analyzed via LCMS.

For isolation of the lasso peptides, 6 L of M9 minimal medium was fermented as described for 3 days at 20 °C and cells were harvested by centrifugation. The cell pellet of each large-scale fermentation was extracted with 400 mL of MeOH and the solvent was evaporated at 40 °C and reduced pressure. Dried extracts were resuspended in 10 mL of 50% MeOH, cleared by centrifugation and filtration and were then subjected to preparative HPLC.

Purification of Caulosegnins I–III. Lasso peptides were purified by preparative HPLC using a microbore 1100 HPLC system (Agilent) with a VP 250/21 Nucleodur C18 Htec 5 μm column (Macherey-Nagel). For detection, the UV absorption at 215 nm was recorded.

Crude pellet extracts were subjected to the following gradient of water/0.045% formic acid (solvent A) and MeOH/0.05% formic acid (solvent B) at room temperature and a flow rate of 18 mL/min: Linear increase from 40% to 85% B in 30 min, followed by a linear increase from 85% to 95% B in 5 min and holding 95% B for additional 3 min. Retention time of caulosegnin I was 17.5 min, retention time of caulosegnin II was 18.0 min and retention time of caulosegnin III was 23.5 min. Lasso peptide containing fractions were pooled and solvent was evaporated at 40 °C and reduced pressure.

For the second round of purification, 1–3 mg of semipure lasso peptide was dissolved in 10 mL of 20% MeCN and subjected to a gradient of water/0.1% trifluoroacetic acid (solvent C) and MeOH/0.05% trifluoroacetic acid (solvent D) at room temperature and a flow rate of 18 mL/min. For purification of caulosegnin I, the following gradient was used: Linear increase from 30% to 45% D in 30 min, followed by a linear increase from 45% to 95% D in 2 min and holding 95% B for additional 3 min. Retention time of caulosegnin I was 12.5 min. For purification of caulosegnin II and III, another gradient was used: Linear increase from 20% to 60% D in 60 min, followed by a linear increase from 60% to 95% D in 5 min and holding 95% B for additional 3 min. Retention time of caulosegnin II was 27.5 and 29.0 min for caulosegnin III. After the second purification, lasso peptide containing fractions were pooled and solvent was evaporated via lyophillization to yield pure product.

NMR Spectroscopy. Samples for NMR measurements contained 4.97 mg of caulosegnin I in 250 μL of methanol-d_3 or the same amount of caulosegnin I in the same volume of methanol-d_4 in Wilmad 3 mm tubes (Rototec Spintec). Spectra were recorded on a Bruker Avance 600 MHz spectrometer equipped with an inverse triple resonance $^{1}H-^{13}C-^{15}N$ probe with z-gradient. Temperature effect on the structure was surveyed by recording ^{1}H spectra at variable temperatures. Thus, all 2D spectra were recorded at 296 and 288 K while structure determination was done on the basis of spectra at 288 K. For sequential assignment, DQF-COSY,[39] TOCSY,[40] NOESY,[41] ROESY,[42] and E.COSY[43] experiments were performed in phase-sensitive mode using States-TPPI.[44] TOCSY spectra were recorded with mixing times of 50 and 80 ms. NOESY spectra were taken at 100, 200, and 300 ms mixing times, while ROESY spectra were observed at 120 ms. Water suppression was fulfilled by using excitation sculpting[45] with gradients for DQF-COSY, TOCSY, NOESY, and ROESY experiments, while presaturation was used for E.COSY. 1D spectra were acquired with 65 536 data points, while 2D spectra were collected using 4096 points in the F_2 dimension and 512 increments in the F_1 dimension. For 2D spectra, 32 transients were used, with an exception of 48 transients for E.COSY. Relaxation delay was 3.0 s. Chemical shifts of ^{1}H were referenced to the rest signal of methanol-d_3/d_4 (δ_{1H} = 3.31 ppm). To study the proton/deuterium exchange, ^{1}H and TOCSY spectra were recorded with the sample in methanol-d_4 at 288 K sequentially 30 min, 1 d, 7 d, 14 d, 21 d and 28 d after sample preparation. All spectra were processed with Bruker TOPSPIN 2.1. NOE cross-peaks were analyzed within the program Sparky.[46]

Structure Calculations. Standard procedures were used.[47] In brief, on the basis of the vicinal coupling constants $^{3}J_{HH\alpha}$ and $^{3}J_{\alpha\beta}$, torsion angles ϕ and χ_1 were determined. NOESY cross-peaks with mixing time 100 ms at 288 K were used to create distance constraints. These constraints were used in the simulated annealing protocol and the structure calculations were done with the program CYANA 2.1.[48] The coordinates of the 15 lowest energy structures to present the solution structure of caulosegnin I have been submitted to the Protein Data Bank (PDB) and assigned the accession number 2LX6.

Thermostability Assays. For testing the thermal stability of the lasso peptides, a solution of 10 μg of purified lasso peptide was incubated at 95 °C for either 1, 2, 4, or 8 h, or for 4 h at either 20, 35, 50, 65, 80, or 95 °C. Samples were cooled to 4 °C and were subsequently analyzed via LCMS using appropriate gradients.

For testing the heat stability of the variants, 50 μL of the respective extracts was incubated at 95 °C for 1 h. The samples were cooled and were analyzed with the LTQ-FT ultra instrument under the above-mentioned conditions. As reference, additional 50 μL of the respective untreated extract was measured under the same conditions.

Protease Assays. To assess the proteolytic stability of caulosegnins I–III, assays were performed using 10 μg of purified lasso peptide and different proteases. Additionally, these assays were performed for the branched cyclic analogues of caulosegnins I–III, which were obtained by heating 10 μg purified lasso peptide to 95 °C for 1 h and were then incubated under the same conditions.

For trypsin, the assays were performed with 0.5 μg of trypsin in a buffer containing 50 mM Tris-HCl and 1 mM $CaCl_2$ at pH 7.6 for 4 h at 25 °C. For chymotrypsin, the assays were performed with 0.5 μg of chymotrypsin in a buffer containing 100 mM Tris-HCl and 10 mM $CaCl_2$ at pH 8.0 for 4 h at 25 °C. For elastase, the assays were performed with 0.5 μg of elastase in a buffer containing 50 mM Tris-HCl at pH 8.5 for 4 h at 37 °C. For proteinase K, the assays were performed with 1 U proteinase K in a buffer containing 50 mM Tris-HCl at pH 7.5 for 4 h at 37 °C. For carboxypeptidase Y, the assays were performed with 0.5 U carboxypeptidase Y in a buffer containing 50 mM MES and 1 mM $CaCl_2$ at pH 6.75 for 4 or 16 h at 25 °C.

Differentiation between Lasso and Branched Cyclic topology. To investigate the topology of a produced compound, several experiments were carried out. The first hint for a lasso peptide topology was derivable from thermal treatment and subsequent LCMS analysis of samples of the putative lasso peptides, where the time and temperature dependent conversion into another topology can be observed by the generation of a new compound with the exact same mass but a different retention time. The putative lasso peptides and their derivatives were then incubated with carboxypeptidase Y to further prove them having either a lasso or a branched cyclic topology. Furthermore, tandem mass spectrometry was applied to gain characteristic fragmentation patterns for both lasso peptides and branched cyclic peptides. For all produced variants discussed in this study, thermal treatment and LCMS analysis combined with tandem mass spectrometry were applied to assess if the isolated variants were of lasso or branched cyclic topology.

Antibacterial Assays. The antibacterial activity of caulosegnin I–III was assayed by spot-on-lawn assays against *Asticaccaulis excentricus*, *Bacillus subtilis*, *Burkholderia thailendensis*, *Burkholderia rhizoxinica*, *Caulobacter crescentus*, *Caulobacter* sp. K31, *C. segnis*, *Micrococcus flavus*, *Sphingobium japonicum*, *Sphingopyxis alaskensis* and *Xanthomonas gardneri*. For this a bacteria, soft agar overlay was prepared by inoculation of 10 mL of an appropriate soft agar medium (6 g/L agar) to an OD of ~0.01 with a culture of the respective bacteria in the same medium during its exponential growth phase. The bacterial suspension was deposited on a 20 mL agar (15 g/L agar) layer of the same medium in a Petri dish. After solidification, droplets containing 0, 5, 10, and 20 nmol of the tested lasso peptide were placed on the agar overlay on marked spots. After 1 and 2 days of incubation at 30 °C, plates were analyzed for the presence of inhibition halos.

LB medium was used for every organism except *C. crescentus* and *C.* sp. K31, which do not grow on LB medium. For these bacteria, the assays were performed with PYEGR medium.

RESULTS AND DISCUSSION

Genome Mining Reveals the Biosynthetic Gene Cluster of the Lasso Peptides Caulosegnins I–III. Through an initial BLAST homology search of the B protein from the capistruin biosynthetic gene cluster, a 648 bp homologue was identified in the genome of *C. segnis* (GenBank accession no. NC_014100) and was designated as *csegB*, according to the general nomenclature for lasso peptide biosynthetic gene clusters.[12,33] Closer inspection of the flanking regions of this gene revealed an 1851 bp gene encoding the asparagine synthetase homologue (*csegC*) as well as three short ORFs with sizes of 111 to 126 bp each. These small peptides were analyzed for the presence of conserved features of lasso peptide precursors, namely a glycine or cysteine residue in the N-terminal region of the peptide, which will be position 1 of the mature lasso peptide, an aspartate or glutamate residue at position 8 or 9 and a threonine residue at the penultimate position of the leader peptide (position −2), which is important for recognition by the processing enzymes.[49] All three peptides met these criteria and their respective genes were named *csegA1*, *csegA2* and *csegA3*. Interestingly CsegA1 features a glutamate capable of macrolactam ring formation at position 8, while CsegA2 and CsegA3 possess suitable glutamate residues at position 9.

Unlike other known lasso peptide biosynthetic gene clusters,[12,13,33,34] that of *C. segnis* does not only feature more than one precursor peptide, but also does not contain a gene encoding an ABC-transporter, which in the cases of MccJ25, capistruin and lariatin were linked with conferring an immunity against the produced compounds.[12,33,34] The latter was also observed for the just recently identified biosynthetic gene cluster of astexin-1,[35] making both systems the first examples of lasso peptide biosynthetic gene clusters arranged in this manner. Interestingly these two systems also share similarities in the genomic regions flanking their gene clusters, as upstream to both clusters a gene encoding a putative GntR family regulator is situated, while downstream a reverse facing gene cluster is positioned, comprising genes proposed to encode a transcriptional regulator, an anti-FecI sigma factor FecR, a TonB-dependent receptor and a peptidase. This is also true for the second putative lasso peptide biosynthetic gene cluster found in the genome of *A. excentricus*, which was predicted by Link and co-workers.[35] Furthermore, it should be noted that even though the gene cluster from *C. segnis* was also identified by the precursor-centric genome mining approach from Link and co-workers, their prediction lacked the precursor peptide from caulosegnin I, showing that a manual quality check for precursor presence and suitability should always be considered.

A scheme of the complete cluster from *C. segnis*, including the precursor sequences, is shown in Figure 1. A comparison of all newly identified genes with their homologues from the MccJ25, astexin-1 and capistruin biosynthetic gene clusters is shown in Table 1.

Fermentation of *C. segnis*. *C. segnis* is a nonpathogenic aquatic alpha proteobacterium of the *Caulobacteraceae* family. This bacterial family is known for their interesting asymmetrical cell division, which is extensively studied using *C. crescentus* as a model organism.[50] Analysis of the *C. segnis* genome with antiSMASH[51] revealed, that it lacks NRPS and PKS gene

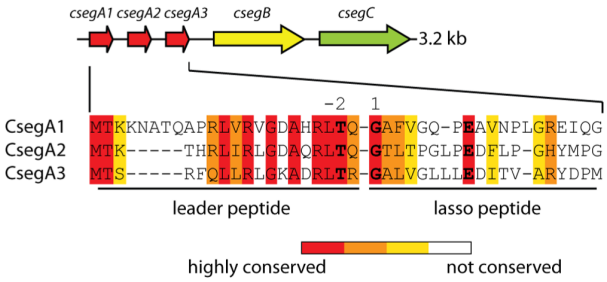


Figure 1. Schematic representation of the caulosegnin biosynthetic gene cluster and alignment of the three precursor peptides.

clusters and thus seems to be confined to the production of secondary metabolites of ribosomal origin like lasso peptides.

After identification of the putative lasso peptide gene cluster of *C. segnis* via genome mining, the strain was cultivated in LB medium, PYEGR medium and several minimal media (M9, M20 and M63) at 20 and 30 °C for 7 days. Best growth was observed in LB and PYEGR medium, while growth in M9 minimal medium was moderate and growth in M20 and M63 minimal media was very poor. Supernatant and pellet extracts were screened for the molecular masses of the predicted lasso peptides by LCMS using a high-resolution Fourier transform mass spectrometer. In the LB, M20 and M63 extracts, no detectable amounts of any lasso peptide were present. In M9 medium, traces of caulosegnin I were found at both 20 and 30 °C in the pellet and supernatant extracts, while PYEGR medium contained traces of caulosegnin I and II in both pellet and supernatant extracts of the 20 °C fermentation and traces of only caulosegnin I in the pellet and supernatant extracts of the 30 °C fermentation (data not shown). But even though it was possible to observe the production of two out of the three predicted lasso peptides, the produced amounts were far too low for isolation and made an heterologous approach necessary.

Heterologous Expression in *E. coli* and Optimization of Caulosegnin Production. For our first efforts to heterologously produce caulosegnins in *E. coli*, the complete 3.2 kb gene cluster was cloned into the pET41a expression vector (Figure 2a).

This construct was expressed in *E. coli* BL21 in LB medium and M9, M20 and M63 minimal media and each fermentation was carried out for either 1 day at 37 °C and 3 days at 20 °C (data not shown). We observed that expression for 3 days at 20 °C in M9 medium was best suited for the production of caulosegnins, while LB was the only medium tested where no lasso peptides could be detected. With this initial system, we were able to detect all three lasso peptides via LCMS, with caulosegnin I showing the best production. Still the production was too low for isolation, which is why single precursor constructs were generated and tested under the same conditions (Figure 2b). Again caulosegnin I showed the best production, which was ~15-fold higher compared to the production in the initial system, while caulosegnin II and III were produced at ~5% of the amount of caulosegnin I. Since the caulosegnin I precursor peptide has a much longer leader sequence compared to the other caulosegnins (Figure 1), we reasoned that this could cause the higher production of caulosegnin I. As consequence hybrid precursor peptides, which carried the leader sequence of CsegA1 and the lasso sequences of CsegA2 and CsegA3, respectively (named CsegA1-QG-A2 and CsegA1-QG-A3), were generated to investigate the effects

Table 1. Coding Sequences from the Caulosegnin Biosynthetic Gene Cluster

protein	amino acids	proposed function	identity/similartiy to MccJ25 cluster homologues[a]	identity/similartiy to capistruin cluster homologues[a]	identity/similartiy to astexin-1 cluster homologues[a]	GenBank accession no.
CsegA1[b]	42	precursor peptide	-	-	-	YP_003593640.1
CsegA2[b]	37	precursor peptide	-	-	-	YP_003593639.1
CsegA3[b]	37	precursor peptide	-	-	-	YP_003593638.1
CsegB	216	protease	20%/35%	24%/34%	28%/43%	YP_003593637.1
CsegC	617	adenylation, cyclization	15%/30%	26%/40%	26%/40%	YP_003593636.1

[a]Determined using the BLOSUM62 algorithm. [b]As precursor peptides are small and as their amino acid sequences are highly variable, identity and similarity was not determined.

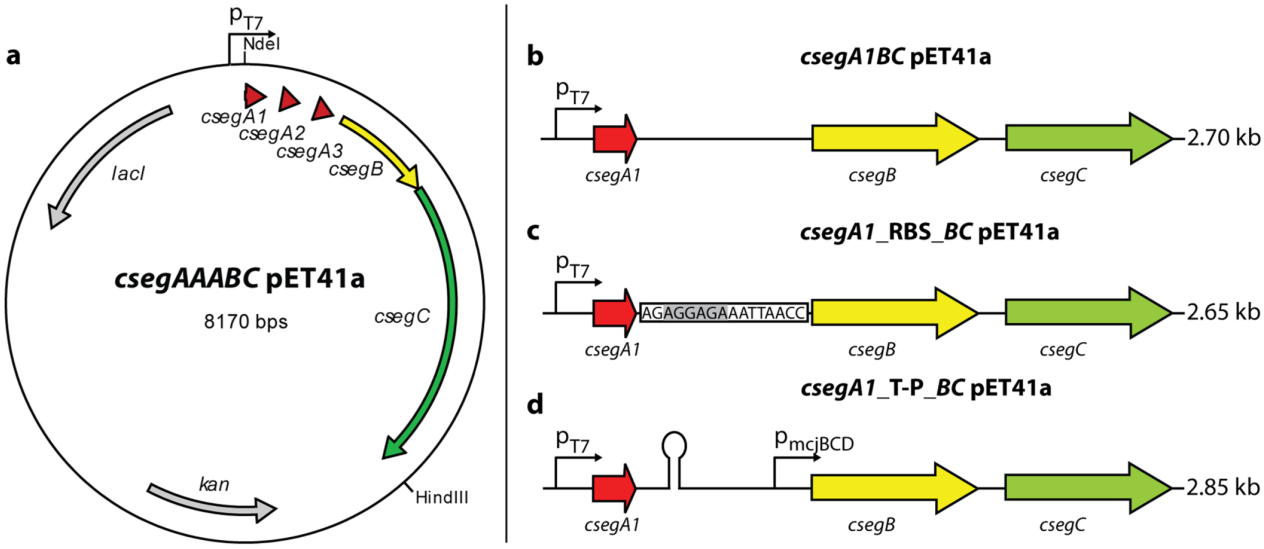


Figure 2. Schematic representation of the different constructs developed for optimized lasso peptide production. (a) Complete construct consisting of the pET41a vector backbone with the inserted *csegAAABC* gene cluster. (b) Gene cluster after deletion of *csegA2* and *csegA3*, which was used as basis for further optimization. (c) Gene cluster after the intergenic region was replaced by an *E. coli* optimized ribosomal binding site (RBS) marked in gray. (d) Gene cluster after the intergenic region was replaced by a terminator sequence (symbolized by a loop) and the subsequent *mcjBCD* promoter.

of the leader peptide on lasso peptide production. This modification led to an improved yield of the lasso peptides, with caulosegnin II now being produced at ~40% and caulosegnin III at ~10% of the amount of caulosegnin I. As caulosegnin I still showed the highest production of all three lasso peptides, it was chosen for additional optimization approaches. In recent publications by Link et al., the production of capistruin,[49] astexin-1[35] and MccJ25[52] was improved through different means. In case of capistruin and astexin-1, production was optimized through exchange of the intergenic region between their respective precursor and processing enzyme genes with an *E. coli* optimized ribosomal binding site. For MccJ25, an improved expression system was generated, where the precursor was under control of the strong inducible T5 promoter, while the expression of the processing enzymes was put under control of their native promoter system.

On the basis of this, one construct (*csegA1*_RBS_*BC* pET41a, Figure 2c) was generated, where the intergenic region between *csegA1* and *csegB* was replaced with the same ribosomal binding site (RBS) and another construct was generated (*csegA1*_T-P_*BC* pET41a, Figure 2d), where this region was replaced with a terminator sequence, followed by the mcjBCD promoter, causing *csegA1* expression to be under control of the strong inducible T7 promoter, while *csegBC* expression levels were now regulated by the mcjBCD promoter system.

The yields with these constructs were vastly improved, being ~11-fold higher for the *csegA1*_RBS_*BC* pET41a and ~16-fold higher for the *csegA1*_T-P_*BC* pET41a construct compared to the *csegA1BC* pET41a construct. As the terminator-promoter (T-P) variant showed the best production, it was also introduced in the *csegA1-QG-A2* and the *csegA1-QG-A3* hybrid systems. The production improvements for caulosegnin II and III were even more dramatic with a ~24-fold increase for caulosegnin II and a ~50-fold increase for caulosegnin III. A comparison of the production of all constructs is shown in Figure 3.

Isolation of Caulosegnins I–III. With the optimized heterologous production of caulosegnins I–III, we performed large-scale fermentations to isolate these lasso peptides. The observed production was ~0.30 mg/L for caulosegnin I, ~0.15 mg/L for caulosegnin II and ~0.10 mg/L for caulosegnin III after two HPLC purification runs (Supporting Information Figure S1–S3) and in accordance with the ratios obtained in the test fermentations. As caulosegnin II and III contain one

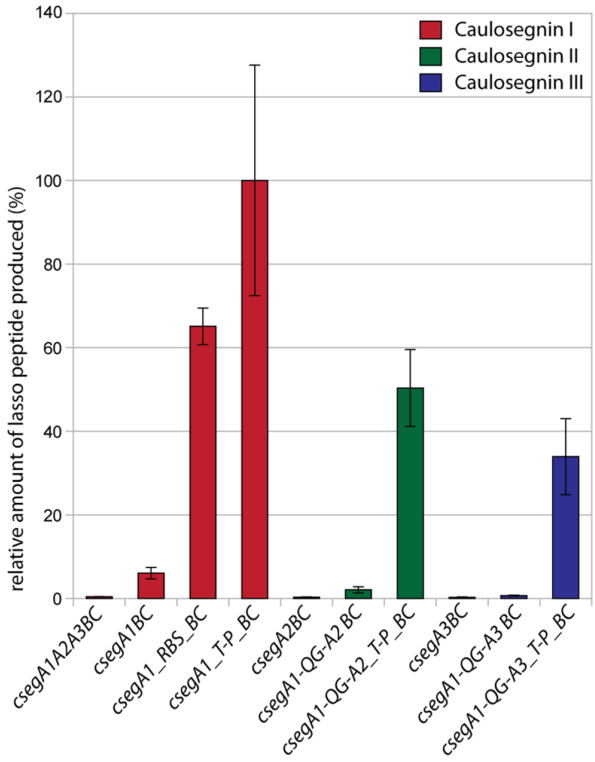


Figure 3. Comparison of relative lasso peptide production for the different pET41a constructs.

methionine residue each, oxidation of these lasso peptides occurs to a small extent when removing the solvent after HPLC purification at 40 °C at reduced pressure. Even though caulosegnin I does not contain any methionine residues, solvent evaporation at 40 °C and reduced pressure led to the formation of small amounts of another compound with exactly the same mass, but a different retention time in the LCMS analysis. This can be explained by the fact that caulosegnin I tends to unthread at higher temperatures.

To prevent the formation of oxidated and unthreaded lasso peptide derivatives, the solvent was removed by lyophillization directly after the final HPLC purification step. This way, pure samples of each lasso peptide could be obtained.

NMR Studies and Structure Elucidation of Caulosegnin I. To gain a better understanding of the structural factors contributing to the unique lasso fold, the structure of the most abundant lasso peptide, caulosegnin I, was elucidated via NMR.[53] Assignments of the ^{1}H signals were obtained by standard procedures.[47] A combination of DQF-COSY and NOESY produced sequential assignments (i.e., all αH and NH and their sequence in the backbone), and a combination of DQF-COSY and TOCSY resulted in assignments of the side chains. One pure conformation was observed. Full assignments of ^{1}H signals were thus obtained (Supporting Information Figure S4–S7, Tables S9–S10). Strong NOE contacts between the NH of Gly1 and the γH of Glu8, showing an internal linkage between the two residues, were observed. Strong sequential αH–δH NOESY cross peaks were detected, which showed the trans-conformation of Pro7 and Pro12. Inspection of ^{1}H spectra between 278 and 298 K (Supporting Information Figure S8) revealed almost no temperature dependence of the NHs of Ala2, Val4, Gln6, Glu8 (positioned in the macrolactam ring), Arg15 and Glu16 (positioned in the tail region and involved in the tail threading/trapping). Variable delay ^{1}H spectra in methanol-d_4 (Supporting Information Figure S9) were consistent with a very slow deuterium exchange of the amide protons of these residues. Furthermore, a large number of long-range NOE contacts were observed. These are the $d_{NN}(i,j)$ connectivities between Ala2–Arg15 and Gln6–Glu16; the $d_{\alpha N}(i,j)$ connectivities between Phe3–Arg15, Pro7–Glu16, and Arg15–Glu8; the $d_{\beta N}(i,j)$ connectivities between Arg15–Val4, Glu16–Gly1, and Glu16–Ala2; and the $d_{\alpha\beta}(i,j)$ connectivity between Val4–Arg15. In addition, contacts between the NH of Gly5 and the δH of Arg15 and between the βH of Phe3 and the εH of Arg15 were also observed. As summarized in Figure 4, all these short distances revealed in the NOESY spectrum are in favor of the lasso structure of caulosegnin I.

Structure calculations were performed with the program CYANA 2.1.[48] The internal linkage was realized by setting the distance constraints between N of Gly1 and Cδ of Glu8 to be 1.33 Å. NOE cross-peaks observed in the 100 ms mixing time NOESY experiment were converted into distance constraints manually. In this way, 148 unambiguous distance constraints were obtained, 43 for the backbone, 28 for long-range interactions, and 77 for the side-chains. Thus, there were an average of 7.8 distance constraints per residue. In addition, constraints of torsion angles ϕ and χ^1 were determined by analyzing the vicinal coupling constants $^3J_{HN\alpha}$ and $^3J_{\alpha\beta}$ (see Supporting Information).

The above-mentioned constraints were used in the simulated annealing protocol for the calculation in the CYANA 2.1 program. The calculation initiated with 50 random conformers and the resultant structures were engineered by the program package Sybyl 7.3[54] to include the covalent linkage between the nitrogen of Gly1 and Cδ of Glu8, followed by energy minimization under NMR constraints using TRIPOS force field within Sybyl. Thus, on the basis of low energies and minimal violations of the experimental data, a family of 15 structures was chosen. These 15 energy-minimized conformers show an average root-mean-square deviation (rmsd) of 0.02 Å and are kept to represent the solution structure of caulosegnin I.

The structure is in accordance with the notion of caulosegnin I being a lasso peptide with an eight amino acid macrolactam ring. The superimposition of the 15 lowest energy structures (Figure 5) shows the backbone adopting a very rigid fold, as the structures show very little deviation from each other, with the last three amino acids below the ring being the most flexible part of the whole lasso peptide.

In regard to the tail passing the ring, Arg15 is located above and Glu16 is the first residue below the ring, suggesting that these amino acids are sterically locking the tail in position (Figure 6). Interestingly Arg15 and Glu16 as well as the C-terminal Gly19 are the only charged residues in caulosegnin I.

It is worth mentioning that the bend caused by the Pro12 residue is the major turning point of the loop and thus Pro12 could play an important role in the prefolding of the precursor peptide and the stabilization of the lasso fold.

Antibacterial Activity Assays. As it is known, for other lasso peptides to exhibit antibacterial activity against related species (e.g., MccJ25[2,28–30] and capistruin[12,31] are able to inhibit the RNA polymerase of some Gram-negative bacteria, but not those of Gram-positive bacteria), caulosegnins I–III were tested primarily for possible antibacterial activity by spot-

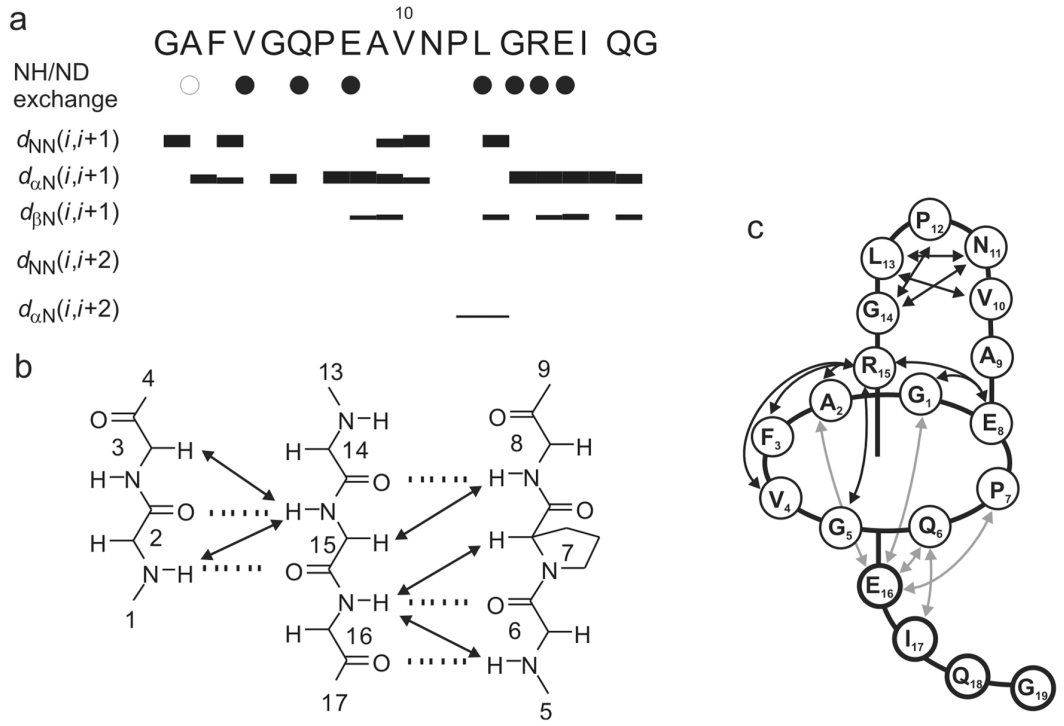


Figure 4. Summary of the NMR data that define the lasso structure of caulosegnin I. (a) The slow (observable in a TOCSY spectrum after 2 days in methanol-d_4 at 288 K) and very slow (observable in a TOCSY spectrum after 4 weeks in methanol-d_4 at 288 K) exchanging amide protons are portrayed with open and filled circles respectively and the sequential and medium range NOE connectivities (observed in a NOESY spectrum of 100 ms mixing time at 288 K in methanol-d_3) are depicted with bars. The thickness of the bars corresponds to the NOE strength. (b) Key backbone−backbone NOE connectivities (100 ms 288 K in methanol-d_3) shown through arrows and hydrogen bond interactions deduced from the amide proton/deuterium exchange experiments and structure calculations shown as broken lines. (c) Schematic view of the caulosegnin I structure showing the long-range NOE connectivities (100 ms 288 K in methanol-d_3) defining the lasso threading. Long-range NOE connectivities in the loop region are shown as black straight arrows and those above and below the ring are shown as black and gray curved arrows respectively.

on-lawn assays against other proteobacteria, namely, *A. excentricus*, *B. thailendensis*, *B. rhizoxinica*, *C. crescentus*, *Caulobacter* sp. K31, *C. segnis*, *S. japonicum*, *S. alaskensis* and *X. gardneri*. Additionally, antibacterial activity was assessed against the Gram-positive bacteria *B. subtilis* and *M. flavus*. In these assays, caulosegnins I−III showed no antibacterial activity against any of the tested strains. An explanation for this could be the lack of an ABC-transporter in the caulosegnin biosynthetic gene cluster, which suggests that these lasso peptides could serve other purposes than being antibacterial agents.

Heat Stability Assays Illustrate the Unique Behavior of Each Caulosegnin. Lasso peptides are often reported to possess an extraordinary stability against thermal degradation.[2,12,22] To investigate if this is also true for the caulosegnins, they were incubated for different time intervals at 95 °C and at different temperatures for 4 h (Supporting Information Figure S11). Interestingly each caulosegnin showed a different stability profile.

While caulosegnin II showed only minimal unthreading even after 8 h at 95 °C, caulosegnin I was completely unthreaded after 4 h at 95 °C. Additionally, at temperatures above 35 and below 95 °C, caulosegnin I shows a strong tendency to form a deamidated derivative, where the carboxamide function of either Gln6, Asn11 or Gln18 is hydrolyzed to a carboxylic acid, resulting in a weight loss of 1 Da and a slight shift of retention time.[55] Caulosegnin III, on the other hand, does not only unthread when exposed to increased temperatures, but also readily decomposes, resulting in small detectable amounts of a derivative missing the last two amino acids and a degradation product which consists of only the nine amino acid macrolactam ring and one additional amino acid. It is worth mentioning that both degradation products feature a C-terminal aspartate, which is in accordance with the known fact that an aspartate residue is able to catalyze the self-hydrolysis of a peptide via formation of a cyclic imide intermediate.[56] As a full-length branched cyclic peptide analogue of caulosegnin III can also be detected, degradation does not seem to be required before the unthreading of caulosegnin III can occur.

For caulosegnin II and III, these assays were performed with samples containing small amounts of oxidized lasso peptide (~15%) to see if the oxidized forms would behave differently. For both caulosegnins, the oxidized forms showed a similar behavior compared to the unmodified lasso peptides.

Different protease assays were performed to compare the stability of the lasso peptides and their branched cyclic analogues. The results of these assays were in agreement with previous observations[22] that lasso peptides are far less susceptible to proteolytic digestion than branched cyclic peptides with the same amino acid sequence (Supporting Information). In this regard, the results from the carboxypeptidase Y digests were most notable, as they allowed a clear distinction between the lasso peptides and their branched cyclic analogues. While the latter were readily digested, the lasso peptides were still intact after protease treatment. Subsequent

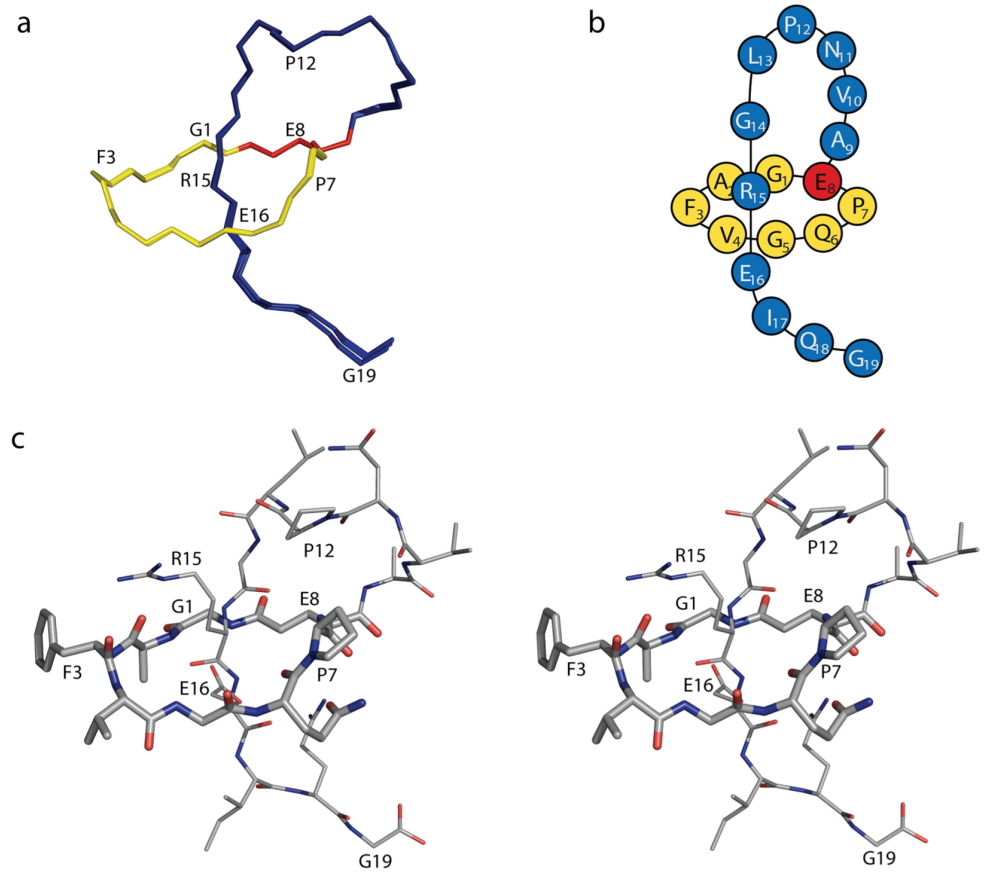


Figure 5. Structure of caulosegnin I. (a) Superimposition of the 15 lowest energy structures of caulosegnin I. Structures are superimposed over all backbone atoms. The isopeptide bond forming Glu8 is colored in red, while the remaining ring amino acids are colored in yellow and the tail amino acids are colored in blue. (b) Schematic representation of the structure of caulosegnin I. (c) Representative average structure of caulosegnin I colored by elements (nitrogen in blue, oxygen in red and carbon in gray) shown in stereoview.

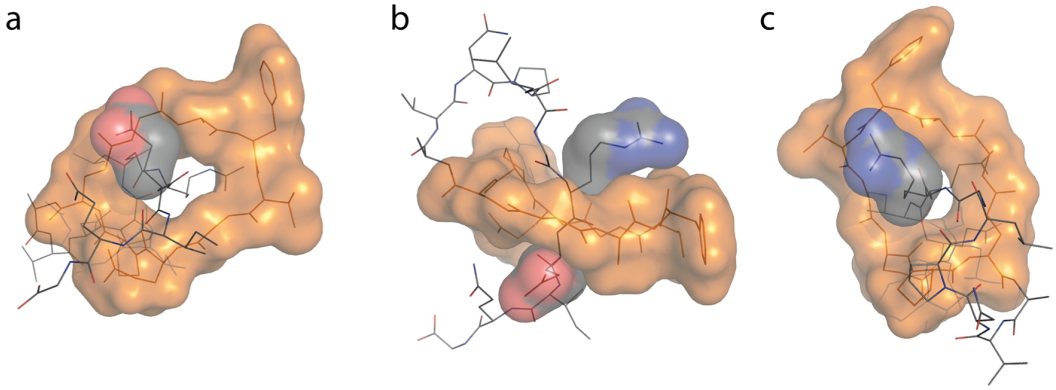


Figure 6. Interactions between the C-terminal tail and the macrocyclic ring of caulosegnin I. The surface of the ring is colored in orange, the surface of the side chains from Arg15 and Glu16 are colored by elements. Steric hindrance caused by the depicted side chains traps the tail within the ring. (a) Shown from below, (b) from the side and (c) from above the ring. A version of this figure in stereoview can be found in the Supporting Information (Figure S10).

MS^2 analysis of the lasso peptides and their branched cyclic analogues revealed distinct fragmentation patterns for the different topologies, where all three lasso peptides differed in similar ways from their branched cyclic analogues (Supporting Information pages S3–S4, Figures S15–S17). As the lasso topology of caulosegnin I was confirmed by NMR structure elucidation, the common features of all three proposed lasso peptides and their branched cyclic analogues are a very strong indication that both caulosegnin II and III are true lasso peptides, although their 3D structures were not yet determined.

Mutational Analysis of Caulosegnin I. In the case of the caulosegnins, the presence of three lasso peptides with different

primary structures is an indication for the high promiscuity of the processing enzymes. This observation is not uncommon for lasso peptide processing enzymes as mutational studies of MccJ25 and capistruin have also shown a very relaxed specificity toward amino acid substitutions in their respective lasso peptide sequences.[22,49,57−59] To further examine which residues are important for efficient processing of the caulosegnin I precursor, a total of 23 mutants was generated (Table 2). To assess if toxicity of a new variant was the reason for a low production, OD_{600} was measured at the end of each fermentation and compared to the OD_{600} of a wild type fermentation. As all OD_{600} values obtained (data not shown) were in the same range, it was shown that the production of caulosegnin I and all its variants had no toxic effects on the cells.

Table 2. Production of the Generated Variants of Caulosegnin I[a]

	relative production	topology	thermal stability
T-2A	*	lasso	no
T-2C	***	lasso	no
T-2I	**	lasso	no
T-2S	**	lasso	no
T-2V	**	lasso	no
G1A	*	lasso	no
G1C	*	lasso	no
G1F	-	-	-
E8D	-	-	-
P12A	**	lasso	no
ΔL13	**	lasso	no
R15A	***	lasso	no
E16A	**	branched cyclic	-
E16D	****	lasso	no
E16K	***	lasso	no
E16L	**	lasso	no
E16N	**	lasso	no
E16Q	***	lasso	no
E16A/I17A	**	branched cyclic	-
I17A	****	lasso	no
I17STOP	*	lasso	no
Q18STOP	***	lasso	no
G19STOP	****	lasso	no

[a]Produced amounts were determined via integration of the corresponding UV signals and comparing them to the wild type lasso peptide. The variants are classified according to their relative production with ***** for >120%, **** for 120−70%, *** for 70−15%, ** for 15−1% and * for <1% of the amount of wild type lasso peptide being produced. Variants that were not produced are marked with "-". For the truncation variants I17STOP, Q18STOP and G19STOP the corresponding amino acid codon was replaced with the TAA stop codon.

The mutational analysis shows that the processing enzymes of the caulosegnin biosynthetic gene cluster tolerate substitutions of the threonine residue at the penultimate position of the leader peptide, but to a lesser extent than the processing enzymes of capistruin and MccJ25.[49] In agreement with the studies for capistruin and MccJ25, the T-2A variant is produced only in trace amounts, while T-2I, T-2S and T-2V are produced by an order of magnitude less than wild type. The best tolerated substitution was T-2C, showing a production that was at least in the same order of magnitude as the wild type.

In accordance with other mutational studies of capistruin[22] and MccJ25,[57] the substitution of Gly1 proved to massively hinder the maturation of caulosegnin I, but unlike capistruin and MccJ25 the G1A and G1C mutations did not completely abolish production as trace amounts of the predicted lasso peptides were still detectable, while the G1F variant did not show any production at all. This is interesting considering that all class I lasso peptides contain a cysteine at position 1, as it again illustrates the different specificity of class I and class II processing enzymes concerning this residue. The E8D variant was also not produced, suggesting that not only the charge of the side chain, but also its length is important for enzymatic recognition. This is consistent with results obtained by the capistruin[22] and MccJ25[57] studies.

The truncation variants I17STOP, Q18STOP and G19STOP provided further information about the enzymatic recognition, as they show that parts of the tail appear to be necessary for a proper maturation of the precursor peptide. While G19STOP is still produced in amounts comparable to wild type, Q18STOP already shows a slight impairment of the production, while the I17STOP variant reduces lasso peptide production to a minimal level.

As expected from the NMR structure, Pro12 seems to play an important role in lasso peptide production. We suggest this is due to the fact that Pro12 induces a bend in the loop region, which could support the prefolding and could facilitate the maintenance of the correct fold during the enzymatic maturation of caulosegnin I.

The R15A mutation was tolerated surprisingly well, indicating that neither a plug above the ring is entirely necessary for the stability of the lasso fold of caulosegnin I, nor that the charge of Arg15 plays a crucial role in the enzymatic recognition. The production of the E16A variant shows, that the charge of the glutamate side chain is also not essential for the enzymatic recognition, but at the same time illustrates that Glu16 is indeed needed to sterically maintain the lasso fold, which apparently can not be sustained by Ile17 or Gln18 as the E16A and E16A/I17A variants could be shown to be branched cyclic peptides. Identification of E16A and E16A/I17A as branched cyclic peptides was possible due to their retention time being much closer to unthreaded than to correctly folded caulosegnin I and due to the fact that retention time was not altered by exposing these variants to increased temperatures (Supporting Information Figure S14), proving that they were already unthreaded.

Furthermore, the systematic exchange of Glu16 with several amino acids revealed that the lasso fold can be maintained by amino acids as small as leucine and asparagine, when they are introduced at position 16 of the lasso peptide. Still the asparagine of the E16N variant is barely able to trap the tail inside of the macrolactam ring, as this lasso peptide is already mostly unthreaded after its isolation. Additionally, the data indicates that an amino acid with a charged side chain at position 16, regardless of being positively or negatively charged, leads to an improved lasso peptide production when compared to amino acids with uncharged side chains.

It is interesting to see that the E16A/I17A mutation results in a branched cyclic peptide, even though the tail features a glutamine residue at position 18, as the E16Q variant was shown to be a lasso peptide. This observation is in accordance with previous reports of capistruin R15A being a heat stable lasso peptide, whose fold is sustained by Phe16, while a R15A/F16A mutation results in a heat labile lasso peptide, whose fold

is still maintained by a phenylalanine, but now at position 18 instead of position 16.[22] In case of both capistruin R15A/F16A and caulosegnin I E16A/I17A the increased flexibility of the tail, caused by the plug being further away from the position where it crosses the macrolactam ring, could explain their altered stability properties even though a suitable plug amino acid is generally present.

Identification of the Plug Amino Acid of Caulosegnin II and III via Mutagenesis. When comparing the amino acid sequences of caulosegnin I with caulosegnin II and III (Figure 7), it is noticeable that each peptide contains a positively

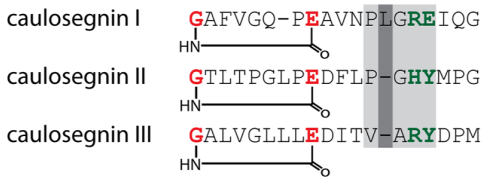


Figure 7. Comparison of the amino acid sequences of caulosegnin I–III. Amino acids involved in the macrolactam ring formation are presented in red. The plug amino acids of caulosegnin I and the proposed plug amino acids of caulosegnin II and III are presented in dark green. Highlighted in light gray are positions which are suggested to have similar importance for the lasso fold. Highlighted in dark gray is the position of Leu13, which in caulosegnin I is inserted in this area.

charged residue in the same distance to the C-terminus, namely Arg15 for caulosegnins I and III, and His15 for caulosegnin II. In analogy to the Glu16 of caulosegnin I, caulosegnin II and III possess a tyrosine residue at position 16, which would be the most suitable plug for the nine amino acid sized ring being present in both of these lasso peptides. Additionally all caulosegnins carry a small amino acid, either glycine or alanine, directly in front of their respective positively charged residue and caulosegnins I and II feature a proline residue four amino acids after their respective macrolactam ring forming glutamate, while caulosegnin III possesses a valine residue at this position.

To identify the plug of caulosegnins II and III, an alanine scan of the last five amino acids of caulosegnin II and III was performed, with exception of the Gly19 residue of caulosegnin II. The results of these mutational studies are shown in Table 3. To assess a possible toxicity of any new lasso peptide variant, OD_{600} was measured and compared to the OD_{600} of the wild type fermentation as described for the mutational analysis of caulosegnin I. Again toxic effects caused by lasso peptide expression were not observed.

As expected, the mutational analysis shows that the Tyr16 residue is important for both caulosegnin II and III. In case of caulosegnin III, the Y16A mutation completely abolishes production, while for caulosegnin II, the Y16A mutation reduces the production to a minimal level of a peptide with a branched cyclic topology. This is in agreement with mutational studies of both capistruin[22] and MccJ25,[57,58] where variants with their lower plug amino acids being replaced with alanine where not produced, while the MccJ25 Y20I variant could only be isolated as a branched-cyclic peptide. The notion of Tyr16 being the plug of both caulosegnin II and III is also supported by results of MS^2 fragmentation studies of the caulosegnins (Supporting Information).

The lack of production of the R15A variant of caulosegnin III highlights that this residue also seems to be very important for either processing, prefolding or stabilization of this lasso

Table 3. Production of the Generated Variants of Caulosegnin II and III[a]

	relative production	topology	thermal stability
caulosegnin II			
P13A	***	lasso	yes
H15A	***	lasso	yes
Y16A	*	branched cyclic	-
M17A	***	lasso	yes
P18A	*****	lasso	yes
caulosegnin III			
V13P	*****	lasso	no
R15A	-	-	-
Y16A	-	-	-
D17A	**	lasso	no
P18A	***	lasso	no
M19A	*****	lasso	no

[a]Produced amounts were determined via integration of the corresponding UV signals and comparing them to the respective wild type lasso peptides. The variants are classified according to their relative production with ***** for >120%, **** for 120–70% *** for 70–15%, ** for 15–1% and * for <1% of the amount of wild type lasso peptide being produced. Variants that were not produced are marked with "-".

peptide. As caulosegnin I R15A and caulosegnin II H15A are produced in rather good amounts, it is most likely that Arg15 is needed to maintain the lasso fold of caulosegnin III, rather than being an important factor for enzymatic recognition and thus for processing of the lasso peptide precursor. This seems plausible since it was already shown that wild type caulosegnin III easily unthreads and it can be assumed that the increased flexibility of the tail through loss of the upper plug would facilitate the unthreading of the lasso peptide even more.

Tyr16 being the plug for both caulosegnin II and III would mean that these lasso peptides must have a loop shortened by one amino acid compared to caulosegnin I, as they feature an one amino acid bigger macrolactam ring. To investigate if this is generally possible, a caulosegnin I ΔL13 variant was generated (Table 2). This variant was successfully processed into a mature lasso peptide, but the strongly reduced production shows that, although this structure is indeed possible for caulosegnin I, a less constraint loop region seems to be preferable for a caulosegnin with an eight amino acid sized macrolactam ring. If this is due to problems during the prefolding step or with the enzymatic recognition remains unclear at this point.

Interestingly, the production of caulosegnin III M19A is improved compared to the wild type. As caulosegnin I and II already feature a small amino acid at the last position of the tail (glycine and alanine respectively) and as the truncation variant of caulosegnin I lacking this last amino acid is produced in amounts comparable to wild type, it could be suggested that a bigger amino acid at this position could lead to a steric repulsion in the enzymatic binding pocket during the lasso peptide maturation and thus decreases the produced amounts of lasso peptide.

As both caulosegnin I and caulosegnin II feature a proline in the same distance to their respective macrolactam ring forming glutamate, a P13A variant of caulosegnin II and a V13P variant of caulosegnin III were generated. These variants further highlight the importance of the proline residue at this position, as the caulosegnin III V13P variant exhibits a strongly increased production, while the produced amounts of caulosegnin II

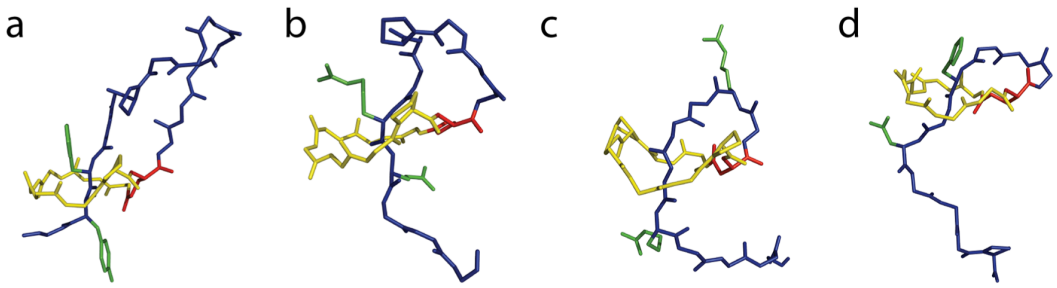


Figure 8. Comparison of class II lasso peptide structures. The ring forming aspartate/glutamate is shown in red, the remaining amino acids of the macrolactam ring are shown in yellow, the tail amino acids are shown in blue and the side chains of the amino acids which were proposed to sterically trap the tail inside the ring are highlighted in green. The shown lasso peptides are MccJ25[18] (a), caulosegnin I (b), capistruin[12] (c) and lariatin A[11] (d).

Table 4. Comparison of Structural Features of Astexin-1, Capistruin, Caulosegnin I, Lariatin A, MccJ25 and RES-701-1

	total length	ring size	tail length	loop length	no. of residues below the ring	ring forming amino acid	upper plug	lower plug
astexin-1	23 aa	9 aa/28 atoms	14 aa	8 aa	6 aa	Asp9	Glu16	Arg18
capistruin	19 aa	9 aa/28 atoms	10 aa	5 aa	5 aa	Asp9	Arg11	Arg15
RES-701-1	16 aa	9 aa/28 atoms	7 aa	4 aa	3 aa	Asp9	Phe12	Tyr14
MccJ25	21 aa	8 aa/26 atoms	13 aa	11 aa	2 aa	Glu8	Phe19	Tyr20
caulosegnin I	19 aa	8 aa/26 atoms	11 aa	7 aa	4 aa	Glu8	Arg15	Glu16
lariatin A	18 aa	8 aa/26 atoms	10 aa	4 aa	6 aa	Glu8	His12	Asn14

P13A decreased compared to its wild type, even though not as drastic as for caulosegnin I P12A. Still it can be assumed that a proline induced bend in the loop could facilitate the prefolding and processing of a lasso peptide precursor and thus could have a positive influence on the produced amounts of lasso peptide.

Comparison of Caulosegnin I to Other Class II Lasso Peptides. There are five other examples of class II lasso peptides with known structures besides caulosegnin I. These are astexin-1,[35] capistruin,[12] lariatin,[11] MccJ25[17−19] and RES-701-1.[20] A comparison of caulosegnin I with capistruin, lariatin and MccJ25 is shown in Figure 8 and Figure S18.

Even though all six lasso peptides belong to the same class, their structures have little in common besides having a lasso fold. All peptides possess differently sized loops and tail sections below the ring that differ in length, with MccJ25 having both the longest loop and shortest tail section below the ring and lariatin A having both the shortest loop and longest tail section below the ring. In regard to the size of the macrolactam ring, astexin-1, capistruin and RES-701-1 are the only ones of these six peptides with a nine membered ring instead of an eight membered one (Table 4). It is noticeable that all these lasso peptides use an aspartate residue for ring formation, while both caulosegnins II and III use a glutamate residue and thus are having a slightly larger macrolactam ring. Analysis of the predicted precursor sequences by Link et al.[35] reveals that the use of an aspartate residue as ring forming amino acid at position 9 appears to be much more common than the use of a glutamate residue at this position. It could be reasoned that through the slightly longer side chain of glutamate, a stabilization of a nine amino acid macrolactam ring could be more difficult, making the use of an aspartate residue favorable as it could lead to a more stable lasso fold.

Interestingly, the amino acids used as plugs are rather diverse, ranging from tyrosine to a comparably small amino acid like asparagine. In general, it appears more and more likely that the size of the amino acid itself is not the only important factor in regard to lasso fold maintenance. In contrast, the size of the loop and the resulting level of strain of its conformation, as well as the distance of the plugs to the part of the tail that passes the ring, could play a major role in stabilizing the lasso fold. This assumption would explain why tyrosine is not capable of stabilizing a 9 aa ring variant of MccJ25,[58] while it was shown that it most likely stabilizes both caulosegnins II and III. The fact that a R15A variant of capistruin is a heat stable lasso peptide, whose fold is sustained by its Phe16 residue, while a R15A/F16A mutation results in a heat labile variant, even though the stabilizing amino acid is still a phenylalanine residue (but now at position 18 instead of 16), further backs this notion. In this case, the elongated tail area that fits through the ring and the corresponding increase of flexibility apparently allow that even a bulky amino acid like phenylalanine may pass through the macrolactam ring.[22]

Taking all this information into consideration, it seems necessary to update the former assumption that only bulky amino acids are capable of maintaining the lasso fold and that the scope of feasible structures of class II lasso peptides could be much more diverse than previously anticipated. As factors like loop length, conformational constraints and orientation of side chains of ring amino acids can hardly be predicted solely based on the amino acid sequence, future genome mining approaches should be realized with an open mind in regard to amino acids capable of being a plug, as even smaller amino acids could be important for the lasso fold.

■ CONCLUSION

In this work, we demonstrate the practical application of the combination of our genome mining approach with an optimized system for heterologous expression in *E. coli* for the isolation of novel lasso peptides. With this approach, we were able to identify a lasso peptide biosynthetic gene cluster within the genome of *C. segnis* and were able to isolate three lasso peptides (caulosegnins I−III). This is the first report since the isolation of capistruin[12] and astexin-1[35] that a new lasso peptide biosynthetic gene cluster was systematically identified

and proven to produce previously unknown lasso peptides. At the same time, this lasso peptide biosynthetic gene cluster is the first instance, together with the cluster from *A. excentricus*, where no immunity conferring ABC-transporter is present.

It is interesting to see that the enzymes of this gene cluster are capable of processing lasso peptides with different ring sizes. Although recent studies showed that it is possible to expand the ring size of MccJ25 through insertion of an alanine residue between the fifth and sixth amino acid of the ring, the result was a mere branched cyclic peptide with a nine membered ring,[58] showing that either a flawed prefolding occurred or that the bulky side chains were no longer able to trap the tail inside the expanded ring. Still the caulosegnin biosynthetic gene cluster is the first known native system to feature not only more than one lasso peptide, but is also the first system being able to process lasso peptides with differently sized rings. Additionally, the production of a lasso peptide featuring a nine amino acid macrolactam ring using as ring forming amino acid a glutamate instead of an aspartate is something not shown before.

The caulosegnins were exposed to temperatures up to 95 °C to test their thermal stability, showing caulosegnin I to unthread, caulosegnin III to unthread and decompose and caulosegnin II to be very stable at increased temperatures. The thermal lability of both caulosegnin I and III is in contrast to the former assumption that lasso peptides have an intrinsic heat stability and raises new questions about factors affecting lasso peptide stabilization. Subsequent assessment of the proteolytic stability of these lasso peptides once more revealed the high resistance of the lasso fold against proteolytic degradation, while the thermally unthreaded analogues of caulosegnin I and III were readily digested by the tested enzymes. Furthermore the structure of caulosegnin I was determined via NMR spectroscopic methods, which showed that a surprisingly small amino acid, Glu16, was maintaining the lasso fold through sterical means.

Via mutational analysis it was shown that the plugs for both caulosegin II and III are most likely their respective Tyr16 residues. To confirm or disprove this assumption, structure elucidation of both lasso peptides is essential and will be possible after their production is further improved to allow isolation of sufficient amounts for NMR spectroscopic analysis. Until then, it will also remain unclear why these peptides show such a different behavior toward increased temperatures, as both feature a nine membered macrolactam ring and possibly use the same amino acid below the ring as plug.

In regard to the biosynthesis of the caulosegnins, our mutational studies revealed that enzymatic recognition most likely does not occur via ionic interactions and that a tail section below the ring of at least three amino acids is needed for efficient processing, while it was also shown that the loop size can be varied between 6 and 7 amino acids.

Our work also shows that the lasso fold of caulosegnin I can be maintained by an amino acid as small as asparagine or leucine, questioning the former notion that only amino acids with big, bulky side chains are capable of fulfilling this function and shifting the paradigm to loop length, structural rigidity and the orientation of the side chains of other amino acids in spacial proximity as additional important factors for lasso peptide stability. The shown importance of a proline residue in the loop region of the caulosegnins further supports this idea, as does the fact that an E16Q variant of caulosegnin I results in a lasso peptide, while an E16A/I17A variant exhibits a branched cyclic topology, even though a glutamine residue is present at position 18 of the tail.

Additionally, our mutational analysis confirmed the importance of the threonine residue at the penultimate position of the leader peptide and the glycine at position 1 of the lasso peptide, which further highlights that these criteria appear to be universal for the maturation of class II lasso peptides. Still we could also show for the first time that the substitution of the glycine at position 1 is not completely unfeasible.

As our antibacterial assays all turned out negative, the native function of these lasso peptides remains elusive. If an antibacterial activity is found in future works, it will further solidify the belief of lasso peptides generally being antibacterial agents giving a selective advantage to its producer under certain conditions. Still, the lack of an immunity conferring ABC-transporter gives rise to the idea of the caulosegnins serving a different function, suggesting that lasso peptides could have a far more diverse scope of biological functions than previously expected.

All these questions illustrate the importance of the systematic discovery, isolation and characterization of new lasso peptides and will lead to a better understanding of their role in nature and the physicochemical factors stabilizing these fascinating molecules.

■ ASSOCIATED CONTENT

Supporting Information

Protease stability tests of the caulosegnins and their branched cyclic analogues, MS^2 studies of the caulosegnins and their branched cyclic analogues, chromatograms of the purifications and of the thermostability assays, DQF-COSY-, TOCSY-, NOESY- and $^{1}H/^{15}N$-HSQC-spectra, ^{1}H spectra at variable temperatures, ^{1}H variable delay spectra, tables showing all used primers, relative production of the different generated constructs and variants, structural statistics of the 15 structures selected to represent the solution structure of caulosegnin I, assignment of ^{1}H and ^{15}N signals and stereoplots of Figures 5 and 7. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

marahiel@staff.uni-marburg.de

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

Financial support from the Deutsche Forschungsgemeinschaft and the LOEWE program of the State of Hesse is gratefully acknowledged. We would like to thank Prof. A. James Link from the University of Princeton for providing us the pJP3 plasmid.

■ REFERENCES

(1) Weber, W.; Fischli, W.; Hochuli, E.; Kupfer, E.; Weibel, E. K. *J. Antibiot.* **1991**, *44*, 164−171.
(2) Salomón, R. A.; Farías, R. N. *J. Bacteriol.* **1992**, *174*, 7428−7435.
(3) Potterat, O.; Stephan, H.; Metzger, J. W.; Gnau, V.; Zähner, H.; Jung, G. *Liebigs Ann. Chem.* **1994**, 741−743.
(4) Morishita, Y.; Chiba, S.; Tsukuda, E.; Tanaka, T.; Ogawa, T.; Yamasaki, M.; Yoshida, M.; Kawamoto, I.; Matsuda, Y. *J. Antibiot.* **1994**, *47*, 269−275.

(5) Ogawa, T.; Ochiai, K.; Tanaka, T.; Tsukuda, E.; Chiba, S.; Yano, K.; Yamasaki, M.; Yoshida, M.; Matsuda, Y. *J. Antibiot.* **1995**, *48*, 1213−1220.
(6) Fréchet, D.; Guitton, J. D.; Herman, F.; Faucher, D.; Helynck, G.; Monegier du Sorbier, B.; Ridoux, J. P.; James-Surcouf, E.; Vuilhorgne, M. *Biochemistry* **1994**, *33*, 42−50.
(7) Tsunakawa, M.; Hu, S. L.; Hoshino, Y.; Detlefson, D. J.; Hill, S. E.; Furumai, T.; White, R. J.; Nishio, M.; Kawano, K.; Yamamoto, S. *J. Antibiot.* **1995**, *45*, 433−434.
(8) Yano, K.; Toki, S.; Nakanishi, S.; Ochiai, K.; Ando, K.; Yoshida, M.; Matsuda, Y.; Yamasaki, M. *Bioorg. Med. Chem.* **1996**, *4*, 115−120.
(9) Kimura, K.; Kanou, F.; Takahashi, H.; Esumi, Y.; Uramoto, M.; Yoshihama, M. *J. Antibiot.* **1997**, *50*, 373−378.
(10) Potterat, O.; Wagner, K.; Gemmecker, G.; Mack, J.; Puder, C.; Vettermann, R.; Streicher, R. *J. Nat. Prod.* **2004**, *67*, 1528−1531.
(11) Iwatsuki, M.; Tomoda, H.; Uchida, R.; Gouda, H.; Hirono, S.; Omura, S. *J. Am. Chem. Soc.* **2006**, *128*, 7486−7491.
(12) Knappe, T. A.; Linne, U.; Zirah, S.; Rebuffat, S.; Xie, X.; Marahiel, M. A. *J. Am. Chem. Soc.* **2008**, *130*, 11446−11454.
(13) Kersten, R. D.; Yang, Y. L.; Xu, Y.; Cimermancic, P.; Nam, S. J.; Fenical, W.; Fischbach, M. A.; Moore, B. S.; Dorrestein, P. C. *Nat. Chem. Biol.* **2011**, *7*, 794−802.
(14) Knappe, T. A.; Linne, U.; Xie, X.; Marahiel, M. A. *FEBS Lett.* **2010**, *584*, 785−789.
(15) Nar, H.; Schmid, A.; Puder, C.; Potterat, O. *ChemMedChem* **2010**, *5*, 1689−1692.
(16) Constantine, K. L.; Friedrichs, M. S.; Detlefsen, D.; Nishio, M.; Tsunakawa, M.; Furumai, T.; Ohkuma, H.; Oki, T.; Hill, S.; Bruccoleri, R. E.; Lin, P. F.; Mueller, L. *J. Biomol. NMR* **1995**, *5*, 271−286.
(17) Bayro, M. J.; Mukhopadhyay, J.; Swapna, G. V. T.; Huang, J. Y.; Ma, L.-C.; Sineva, E.; Dawson, P. E.; Montelione, G. T.; Ebright, R. H. *J. Am. Chem. Soc.* **2003**, *125*, 12382−12383.
(18) Rosengren, K. J.; Clark, R. J.; Daly, N. L.; Göransson, U.; Jones, A.; Craik, D. J. *J. Am. Chem. Soc.* **2003**, *125*, 12464−12474.
(19) Wilson, K. A.; Kalkum, M.; Ottesen, J.; Yuzenkova, J.; Chait, B. T.; Landick, R.; Muir, T.; Severinov, K.; Darst, S. A. *J. Am. Chem. Soc.* **2003**, *125*, 12475−12783.
(20) Katahira, R.; Shibata, K.; Yamasaki, M.; Matsuda, Y.; Yoshida, M. *Bioorg. Med. Chem.* **1995**, *3*, 1273−1280.
(21) Wyss, D. F.; Lahm, H. W.; Manneberg, M.; Labhart, A. M. *J. Antibiot.* **1991**, *44*, 172−180.
(22) Knappe, T. A.; Linne, U.; Robbel, L.; Marahiel, M. A. *Chem. Biol.* **2009**, *16*, 1290−1298.
(23) Yamasaki, M.; Yano, K.; Yoshida, M.; Matsuda, Y.; Yamaguchi, K. *J. Antibiot.* **1994**, *43* (3), 276−280.
(24) Blond, A.; Cheminant, M.; Destoumieux-Garzón, D.; Ségalas-Milazzo, I.; Goulard, C.; Rebuffat, S. *Eur. J. Biochem.* **2002**, *269*, 6212−6222.
(25) Rosengren, K. J.; Blond, A.; Afonso, C.; Tabet, J. C.; Rebuffat, S.; D., J. *Biochemistry* **2004**, *43*, 4696−4702.
(26) Soudy, R.; Wang, L.; Kaur, K. *Bioorg. Med. Chem.* **2012**, *20*, 1794−1800.
(27) Katahira, R.; Shibata, K.; Yamasaki, M.; Matsuda, Y.; Yoshida, M. *Bioorg. Med. Chem. Lett.* **1995**, *5*, 1595−1600.
(28) Yuzenkova, J.; Delgado, M.; Nechaev, S.; Savalia, D.; Epshtein, V.; Artsimovitch, I.; Mooney, R. A.; Landick, R.; Farias, R. N.; Salomon, R.; Severinov, K. *J. Biol. Chem.* **2002**, *277*, 50867−50875.
(29) Adelman, K.; Yuzenkova, J.; La Porta, A.; Zenkin, N.; Lee, J.; Lis, J. T.; Borukhov, S.; Wang, M. D.; Severinov, K. *Mol. Cell* **2004**, *14*, 753−762.
(30) Semenova, E.; Yuzenkova, Y.; Peduzzi, J.; Rebuffat, S.; Severinov, K. *J. Bacteriol.* **2005**, *187*, 3859−3863.
(31) Kuznedelov, K.; Semenova, E.; Knappe, T. A.; Mukhamedyarov, D.; Srivastava, A.; Chatterjee, S.; Ebright, R. H.; Marahiel, M. A.; Severinov, K. *J. Mol. Biol.* **2011**, *412*, 842−848.
(32) Knappe, T. A.; Manzenrieder, F.; Mas-Moruno, C.; Linne, U.; Sasse, F.; Kessler, H.; Xie, X.; Marahiel, M. A. *Angew. Chem.* **2011**, *50*, 8714−8717.
(33) Solbiati, J. O.; Ciaccio, M.; Farías, R. N.; González-Pastor, J. E.; Moreno, F.; Salomón, R. A. *J. Bacteriol.* **1999**, *181*, 2659−2662.
(34) Inokoshi, J.; Matsuhama, M.; Miyake, M.; Ikeda, H.; Tomoda, H. *Appl. Microbiol. Biotechnol.* **2012**, *95*, 451−460.
(35) Maksimov, M. O.; Pelczer, I.; Link, A. J. *Proc. Natl. Acad. Sci. U.S.A.* **2012**, *109*, 15223−15228.
(36) Yan, K. P.; Zirah, S.; Goulard, C.; Knappe, T. A.; Marahiel, M. A.; Rebuffat, S. *ChemBioChem* **2012**, *13*, 1046−1052.
(37) Duquesne, S.; Destoumieux-Garzón, D.; Zirah, S.; Goulard, C.; Peduzzi, J.; Rebuffat, S. *Chem. Biol.* **2007**, *14*, 793−803.
(38) Pan, S. J.; Rajniak, J.; Cheung, W. L.; Link, A. J. *ChemBioChem* **2012**, *13*, 367−37.
(39) Rance, M.; Sørensen, O. W.; Bodenhausen, G.; Wagner, G.; Ernst, R. R.; Wüthrich, K. *Biochem. Biophys. Res. Commun.* **1983**, *117*, 479−485.
(40) Bax, A.; Davis, D. G. *J. Magn. Reson.* **1985**, *65*, 355−360.
(41) Jeener, J.; Meier, B. H.; Bachmann, P.; Ernst, R. R. *J. Chem. Phys.* **1979**, *71*, 4546−4553.
(42) Bax, A.; Davis, D. G. *J. Magn. Reson.* **1985**, *63*, 207−213.
(43) Griesinger, C.; Sørensen, O. W.; Ernst, R. R. *J. Magn. Reson.* **1987**, *75*, 474−492.
(44) Marion, D.; Ikura, M.; Tschudin, R.; Bax, A. *J. Magn. Reson.* **1989**, *85*, 393−399.
(45) Hwang, T.-L.; Shaka, A. J. *J. Magn. Reson., Ser. A* **1995**, *112*, 275−279.
(46) Goddard, T. D.; Kneller, D. J. *Sparky 3*; University of California: San Francisco.
(47) Wüthrich, K. *NMR of Protein and Nucleic Acids*; Wiley: New York, 1986.
(48) Herrmann, T.; Güntert, P.; Wüthrich, K. *J. Mol. Biol.* **2002**, *319*, 209−227.
(49) Pan, S. J.; Rajniak, J.; Maksimov, M. O.; Link, A. J. *Chem. Commun.* **2012**, *48*, 1880−1882.
(50) Curtis, P. D.; Brun, Y. V. *Microbiol. Mol. Biol. Rev.* **2010**, *74*, 13−41.
(51) Medema, M. H.; Blin, K.; Cimermancic, P.; de Jager, V.; Zakrzewski, P.; Fischbach, M. A.; Weber, T.; Takano, E.; Breitling, R. *Nucleic Acids Res.* **2011**, *39*, W339−3W46.
(52) Pan, S. J.; Cheung, W. L.; Link, A. J. *Protein Expression Purif.* **2010**, *71*, 200−206.
(53) Xie, X.; Marahiel, M. A. *ChemBioChem* **2012**, *13*, 621−625.
(54) *Sybyl 7.3*; Tripos, Inc.: St. Louis, MO.
(55) Radkiewicz, J. L.; Zipse, H.; Clarke, S.; Houk, K. N. *J. Am. Chem. Soc.* **2001**, *123*, 3499−3506.
(56) Skribanek, Z.; Mezo, G.; Mák, M.; Hudecz, F. *J. Pept. Sci.* **2002**, *8*, 398−406.
(57) Pavlova, O.; Mukhopadhyay, J.; Sineva, E.; Ebright, R. H.; Severinov, K. *J. Biol. Chem.* **2008**, *283*, 25589−25595.
(58) Ducasse, R.; Yan, K. P.; Goulard, C.; Blond, A.; Li, Y.; Lescop, E.; Guittet, E.; Rebuffat, S.; Zirah, S. *ChemBioChem* **2012**, *13*, 371−380.
(59) Pan, S. J.; Link, A. J. *J. Am. Chem. Soc.* **2011**, *133*, 5016−5023.

3.2 The Astexin-1 Lasso Peptides: Biosynthesis, Stability, and Structural Studies

Marcel Zimmermann, Julian D. Hegemann, Xiulan Xie, and Mohamed A. Marahiel, The Astexin-1 Lasso Peptides: Biosynthesis, Stability, and Structural Studies,
Chem. Biol. **2013**, *20(4)*, 558-569.
doi: 10.1016/j.chembiol.2013.03.013

Author contributions:

The project was designed by Marcel Zimmermann and Mohamed A. Marahiel with assistance of Julian D. Hegemann. All experiments were carried out by Marcel Zimmermann, except for the structure revision of astexin-1 by NMR spectroscopy which was done by Xiulan Xie. The experimental data was evaluated and interpreted by Marcel Zimmermann with the aid of Julian D. Hegemann. The manuscript was written by Marcel Zimmermann and its revision was done together with Julian D. Hegemann and Mohamed A. Marahiel, of whom the latter was additionally responsible for the project supervision in general.

Background and Summary:

During the initial experimental work for this study, the isolation of the lasso peptide astexin-1 and the elucidation of its three-dimensional structure were published by Link and co-workers.[30] Still, the structure postulated was in strong disagreement with information that we had already obtained from mutational studies, namely that the plug residues proposed by Link and co-workers were easily interchangeable with alanine, while the exchange of Y14 and F15 with alanine dramatically decreased the lasso peptide production. Encouraged by this, we elucidated the three-dimensional structure of a truncated variant of astexin-1 ourselves and indeed could prove that the previously published structure was flawed. We subsequently started to search for reasons for this discrepancy and characterized the stability properties of astexin-1. Besides probing the proteolytic stability of this lasso peptide, the thermal stability of astexin-1 was also assessed, revealing that it was, similarly to the previously published caulosegnins I and III,[22] a heat sensitive lasso peptide. Based on this, it could be reasoned that a too careless purification of astexin-1 in the work done by Link and co-workers led to the unthreading of the lasso fold and that they merely analyzed a branched-cyclic peptide by NMR that was falsely assigned as lasso peptide, which might explain the poor quality of the presented NMR structure. In addition to these results, this study was substantiated by a thorough mutational analysis of the astexin-1 precursor peptide. The most notable result of these investigations was that astexin-1 could be transformed into a heat stable lasso peptide by simple exchange of the lower plug amino acid F15 with the more bulky residue tryptophan.

The Astexin-1 Lasso Peptides: Biosynthesis, Stability, and Structural Studies

Marcel Zimmermann,[1] Julian D. Hegemann,[1] Xiulan Xie,[1,*] and Mohamed A. Marahiel[1,*]
[1]Department of Chemistry, Biochemistry, Philipps-University Marburg, Hans-Meerwein-Strasse 4 and LOEWE-Centre for Synthetic Microbiology, D-35032, Marburg, Germany
*Correspondence: xie@mailer.uni-marburg.de (X.X.), marahiel@staff.uni-marburg.de (M.A.M.)
http://dx.doi.org/10.1016/j.chembiol.2013.03.013

SUMMARY

Lasso peptides are a large family of natural products that owe their name to a unique structure formed by a side chain to backbone macrocyclization, resembling a knotted lasso. The unique structure has significant impact on their biological and physical properties, as lasso peptides are usually more stable than linear ones. Current work examines stability, structure, and biosynthesis of recently discovered lasso peptide astexin-1, a heat-sensitive lasso peptide. The obtained results revealed a new lasso structure with a tight loop and long tail as well as narrow specificity of the maturation machinery for some essential residues associated with the protease processing site, involved in macrolactam ring formation and entrapment of the tail. Using the astexin-1 structure, it was possible to rationally construct a thermostable variant of this lasso peptide.

INTRODUCTION

Ribosomally assembled natural products often compensate for the lack of diversity in building blocks, compared to NRPS- or PKS-derived molecules, with intensive posttranslational modifications ranging from acylation over thiazolation/oxazolation to macrocyclization (Arnison et al., 2013). These modifications modulate their bioactivity and stabilize them against degradation. Lasso peptides are quite an exception in this regard, because their only enzymatically catalyzed posttranslational modification is a side chain to backbone macrocyclization, resulting in a special three-dimensional (3D) structure that resembles a knotted lasso. This unusual fold stabilizes them against heat, chemical denaturation, and proteolytic degradation. The macrolactam ring is formed between the N terminus of a glycine or cysteine at the first position of the core peptide and the acidic side chain of an aspartate or glutamate residue at position 8 or 9. Besides these two highly conserved residues, the primary sequence is in general variable. In the C-terminal tail sequence, usually one or two residues with bulky side chains are present that sterically entrap the tail inside the ring and prevent unfolding (Duquesne et al., 2007).

The known lasso peptides are divided into three groups according to the absence or presence of conserved cysteine residues. Class I lasso peptides contain four cysteine residues, including position 1, that form two disulfide bonds; class II lasso peptides lack cysteine residues; and class III, with its only representative being BI-32169 (Knappe et al., 2010; Nar et al., 2010), features two cysteine residues forming one disulfide bond, whereas position 1 is a glycine as in class II (Xie and Marahiel, 2012).

In addition to their peculiar structure, lasso peptides possess some interesting bioactivities ranging from antiviral activity against HIV (Fréchet et al., 1994; Lin et al., 1996), inhibition of the glucagon receptor (Potterat et al., 2004), the endothelin type B receptor (Katahira et al., 1995; Yano et al., 1995), and the atrial natriuretic factor (Weber et al., 1991; Wyss et al., 1991) to antimicrobial activities (Iwatsuki et al., 2007; Kimura et al., 1997; Yano et al., 1996). Becausee most of these effects were found in focused systematic screens, they may merely be coincidental, whereas the antibacterial activities may likely resemble their native function. For the best-studied lasso peptide, microcin J25, and the genome-mining-derived capistruin, the target has been identified to be the bacterial RNA polymerase (Adelman et al., 2004; Kuznedelov et al., 2011).

Although as few as five lasso peptide biosynthetic gene clusters are known so far, their organization is quite simple as they only consist of three to five different genes. For the prototypical lasso peptide microcin J25, the gene organization is as follows. The *mcjA* gene encodes the 58 residue lasso precursor peptide, which is matured in two steps by the transglutaminase core domain-containing protein McjB that is assumed to cleave off the 37 residue leader peptide and the ATP-dependent asparagine synthetase homolog McjC, which is believed to be involved in the macro cyclization by activation of the acidic side chain. These catalytic steps were recently shown in vitro, and therefore the formerly proposed functions of the gene products could be verified (Pan et al., 2012a; Yan et al., 2012). Furthermore, the transglutaminase homolog McjB, which contains a catalytic triad consisting of cysteine, histidine, and aspartate and therefore acts as a cysteine protease, is also believed to be associated with peptide prefolding, because ATP consumption was essential for catalysis. The maturated lasso peptide is then exported by McjD, an ATP-binding cassette transporter protein. In contrast to the four gene clusters of microcin J25 and capistruin, the recently discovered lariatin biosynthetic gene cluster features an additional gene *larC*. This gene is located upstream of the gene *larD*, which encodes a putative protease. It shows homology to the coenzyme pqq synthesis protein D and was shown to be essential for lariatin production (Inokoshi et al., 2012). The most recently discovered lasso peptides caulosegnins I –III originate from a single cluster featuring three

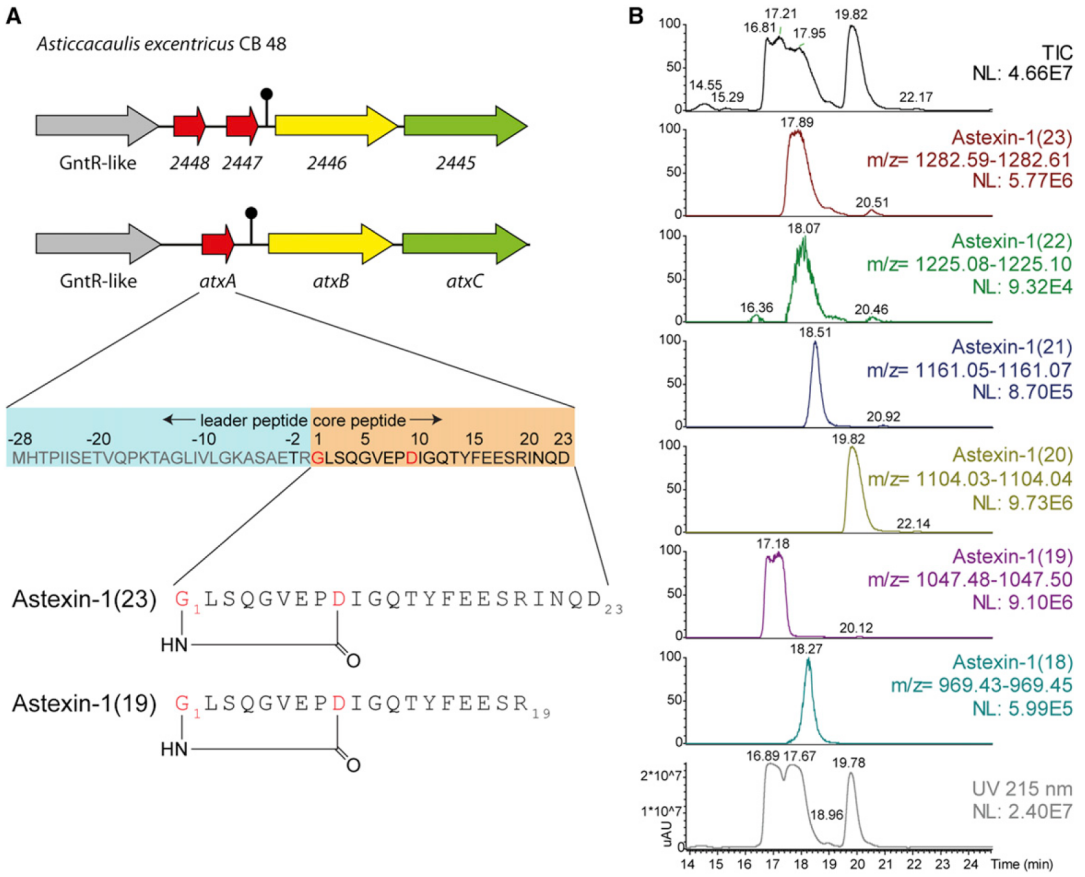


Figure 1. Expression of the Astexin-1 Biosynthetic Gene Cluster

(A) Schematic representations of the two lasso peptide gene clusters of *Asticcacaulis excentricus*: the 244x cluster featuring two potential lasso products astexin-2 and astexin-3 and the astexin-1 gene cluster (222x/*atxABC*). Full sequence of the astexin-1 precursor peptide and the two main astexin-1 products: astexin-1(19) and astexin-1(23).

(B) HPLC-MS analysis of the cell pellet of an *E. coli* strain transformed with Astex_222xABC_pET41 shown as total ion current (TIC) in black; extracted ion chromatograms (EIC) and UV signal shown in gray. Extracted masses fit to astexin-1(23) red, (22) green, (21) blue, (20) ocher, (19) purple, and (18) teal. Additional information regarding the purification of heterologously produced astexin-1 is shown in Figure S1.

See also Table S5.

precursors, while lacking an ABC transport protein (Hegemann et al., 2013).

Numerous putative lasso peptide biosynthetic gene clusters were recently identified in a precursor-driven genome mining approach (Maksimov et al., 2012). Besides the basic constrictions for a lasso peptide precursor (glycine at position 1, glutamate, or aspartate at positions 6 to 10), conserved motifs in the two processing enzymes B and C were defined and used for identification of promising clusters. The biosynthetic gene cluster responsible for the production of astexin-1 was chosen for further investigation, and the structure of the full-length astexin-1(23) was obtained.

In this paper, we present a thorough study on two of the main products of the astexin-1 biosynthetic gene cluster, astexin-1(19) and astexin-1(23), including biosynthesis, mutagenesis, thermal stability, and proteolytic stability of the two main astexins and variants generated by mutagenesis. The results we obtained were inconsistent with the recently published structure (Maksimov et al., 2012). To clarify this apparent contradiction, we reinvestigated the 3D nuclear magnetic resonance (NMR) structure of astexin-1(19) in aqueous solution. The NMR spectroscopic investigation revealed a structure for this lasso peptide that is in accordance with all results of the stability and mutational analysis.

RESULTS AND DISCUSSION

The Astexin Biosynthesis Cluster

Using a general genome mining approach established by our group, we were able to identify two putative lasso peptide biosynthetic gene clusters in the alphaproteobacterium *Asticcacaulis excentricus* CB48 (accession no. NC_014816.1 and NC_014817.1). These clusters were also discovered in the recent publication by Maksimov et al. (2012). One cluster features a single precursor, whereas the other contains two putative precursors. Remarkably, no ABC transporter was found in a vicinity of 10 kb of either cluster, while both clusters contain a GntR-type regulator upstream of the precursor gene (Figure 1A).

Furthermore, analogous regions could be found downstream of both clusters featuring genes encoding a putative peptidase as well as a TonB-dependent receptor. Both clusters were examined for lasso peptide production in fermentations.

Fermentation of *A. excentricus*

In the recent publication covering the discovery of astexin-1, no production of the putative lasso peptide was detectable when *A. excentricus* was grown in peptone yeast extract (PYE) medium (Maksimov et al., 2012). In this study, a more extensive analysis for homologous production was performed for which *A. excentricus* was fermented in PYE, peptone yeast extract glucose riboflavin, lysogeny broth, M9, M20, and M63 mediums at 30°C. Whereas no growth was observed in M20 and M63, pellets and supernatants of the other media were extracted with methanol in accordance with the protocol stated in the Experimental Procedures. The extracts were evaporated, resuspended in 50% methanol, and subsequently analyzed via high-performance liquid chromatography (HPLC) coupled with high-resolution Fourier transformation mass spectrometry (FT-MS) (in total abbreviated HPLC-MS). The resulting spectra were screened for the predicted masses of truncated (Δ1 to Δ7) and full-length products of the single precursor cluster 222x (23 residues) and of the two precursor cluster 244x (24 residues). In the pellet and supernatant extracts of M9 medium, cultures ions of the product of cluster 222x (Figure 1A) in full-length (m/z_{obs} = 1,282.6036 $[M+2H]^{2+}$), termed astexin-1(23), and shortened by four amino acids (Ile20-Asn21-Gln22-Asp23) (m/z_{obs} = 1,047.4944 $[M+2H]^{2+}$), termed astexin-1(19), as well as the first product of cluster 244x (Figure 1A) shortened by three amino acids (Ser22-Ala23-Asp24) (m/z_{obs} = 1,114.5901 $[M+2H]^{2+}$), termed astexin-2(21), were detected (data not shown). As only a low production was observed, a heterologous expression approach was necessary.

Cloning and Expression of the Astexin-1 Lasso Peptide Biosynthesis Cluster

The cluster astex_2228-2230ABC (*atxABC*) (Figure 1A) was cloned into the pET41a expression vector carrying the T7 promoter, introduced into *Escherichia coli* BL21 (DE3) by transformation, and the resulting strain was fermented in M9 minimal medium at various temperatures and incubation times. After methanol extraction of both pellet and supernatant, HPLC-MS analysis revealed masses matching to the predicted masses of 23 amino acid full-length peptide astexin-1(23) and truncated derivatives missing 1 to 5 amino acids, termed astexin-1(22) to astexin-1(18) (Figure 1B). The relative amounts of full-length and truncated variants did not significantly vary upon different fermentation times and temperatures, whereas the total peptide production was best at 37°C overnight. Comparison of UV signal intensities of the three main products, astexin-1(23), (20), and (19), in pellet and supernatant extracts revealed that the amounts in the pellet exceed the amounts in the supernatant by approximately an order of magnitude, which is an interesting finding, as in the recent publication (Maksimov et al., 2012) only trace amounts of the products were found in the pellet.

Furthermore, the Δ5 truncated variant astexin-1(18) is of particular interest, as this compound lacks the amino acid Arg19, which was proposed in the recently published structure (Maksimov et al., 2012) to act as the plug.

Thermal and Proteolytic Stability of Astexin-1

So far, most lasso peptides have been considered to be highly stable against thermal or chemical denaturation and proteolytic digestion because of their inimitable three-dimensional fold. However, only the lasso peptides microcin J25 and capistruin and the caulosegnins have been investigated for thermal stability (Hegemann et al., 2013; Knappe et al., 2009; Salomón and Farías, 1992). Although all lasso peptides are stable at room temperature, microcin J25, capistruin, and caulosegnin II can also withstand prolonged exposure to 95°C, whereas caulosegnins I and III were susceptible to thermal unfolding at high temperatures and are therefore considered thermolabile lasso peptides. This finding encouraged us to study the thermostability of astexin-1. When astexin-1(23) was subjected to 95°C for 1, 2, 4, and 8 hr and analyzed via HPLC-MS, a second peak with the same mass and a significantly different retention time was detectable (Figure 2C). This conversion was irreversible, indicating that it likely represents an unfolding process. Because of the fast conversion, which was almost completed after 1 hr, the peptide was also subjected to lower temperature treatments (20°C, 35°C, 50°C, 65°C, and 80°C) for 4 hr (Figure 2D). Under these conditions, the peptide started unfolding at 50°C. The same overall behavior was also observed for the 19 amino acid (aa) variant astexin-1(19) (Figures 2A and 2B), which lead us to infer that the terminal 4 aa have only little influence on the general heat stability. Interestingly, in the range from 50°C to 65°C, a third peak with the same mass could be observed for astexin-1(23), which may be an intermediate of the unfolding process (Figures 2C and 2D).

Both astexin-1(23) and astexin-1(19) and their unfolded variants were analyzed by tandem mass spectrometry, revealing a very similar fragmentation pattern (Figures 2E–2H). Only minor differences occur in the intensity of certain ions when comparing the spectra of the respective lasso peptides with their branched cyclic analogs. The most prominent ions for the lasso peptides always correspond to the loss of small neutral molecules like water or ammonia, whereas the unfolded variants have smaller fragments as the dominating peaks in the spectra. No binary peptide fragments could be identified due to the low energy of the used fragmentation method (see Experimental Procedures). A MALDI-TOF tandem mass experiment reproducing the conditions used in the recent publication (Maksimov et al., 2012) revealed the presence of mass signals proposed as binary peptide fragments. However, the same mass signals were found in a thermally unfolded sample of astexin-1(23) (Figure 2I), where the occurrence of binary peptide fragments is impossible.

The thermal instability of astexin-1 may also explain why in this recent publication (Maksimov et al., 2012) only trace amounts of astexin-1 were detected in the pellet as the extraction method used in this publication involved boiling the cells.

The lasso peptides lariatin and caulosegnins I–III have also been explored toward their stability against carboxypeptidases (Hegemann et al., 2013; Iwatsuki et al., 2006). These enzymes generally digest amino acids from the C terminus of an unfolded protein or peptide. While the lasso fold protects part of the extracyclic tail, an unfolded, branched cyclic lasso peptide

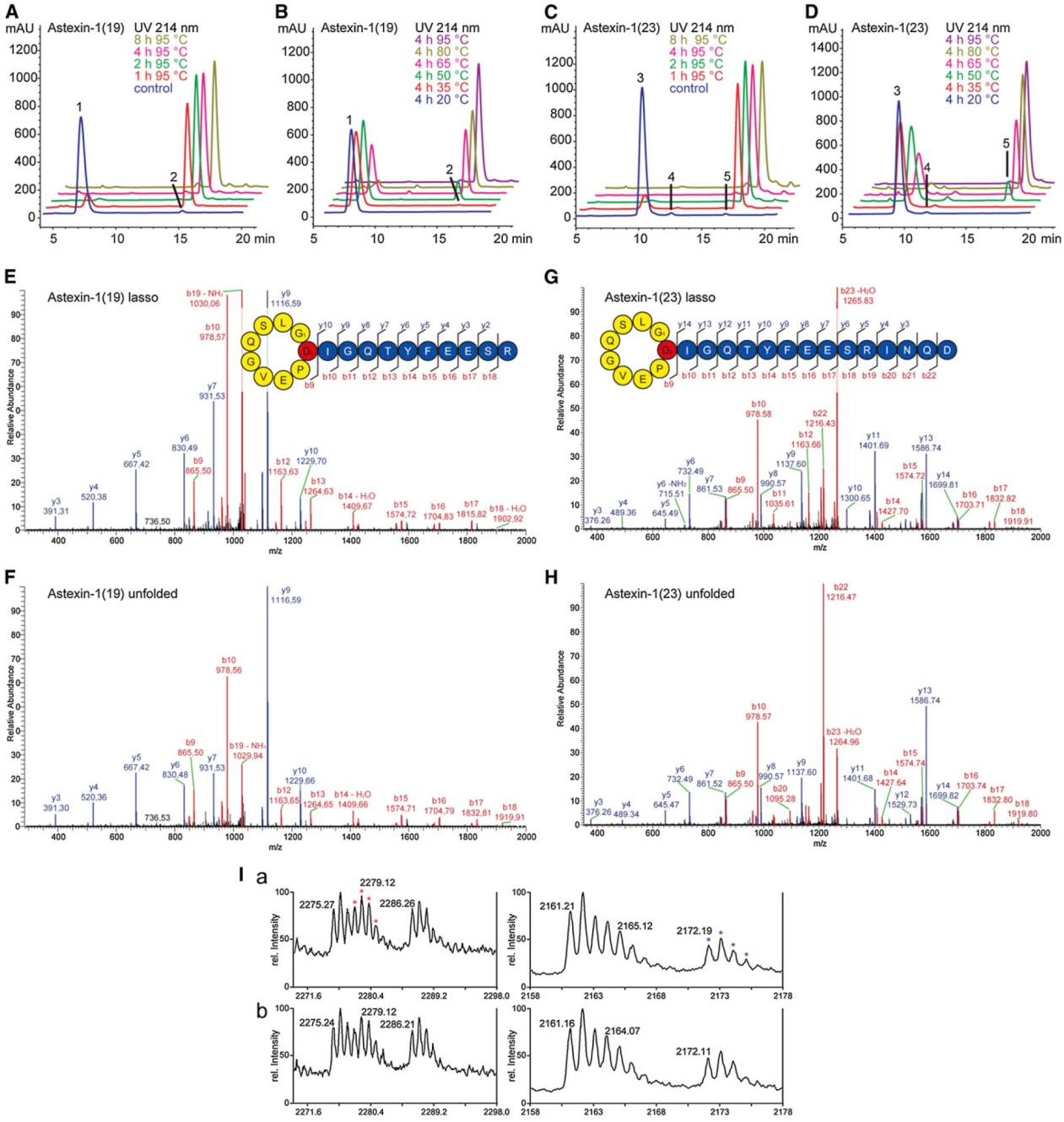


Figure 2. Heat Stability of the Two Major Compounds Astexin-1(19) and (23), Shown as HPLC Chromatograms and Detailed Fragmentation Spectra

(A and B) Astexin-1(19) at (A) 95°C and (B) after heat treatment at lower temperatures. **1** astexin-1(19) lasso, **2** compound with the same m/z ratio as astexin-1(19), carboxypeptidase Y digest proves this to be the unfolded variant.

(C and D) Astexin-1(23) at (C) 95°C and (D) after heat treatment at lower temperatures. **3** astexin-1(23) lasso, **4** compound with the same m/z ratio as astexin-1(23), may be an unfolding intermediate, **5** compound with the same m/z ratio as astexin-1(23), carboxypeptidase Y digest proves this to be the unfolded variant.

(E–H) Fragmentation spectra of (E) astexin-1(19) lasso (compound **1**) and (F) astexin-1(19) unfolded (compound **2**). Fragmentation spectra of (G) astexin-1(23) lasso (compound **3**) and (H) astexin-1(23) unfolded (compound **5**).

(I) MALDI-TOF spectra of astexin-1(23) in **a** the lasso fold and **b** unfolded shown in the same regions as those used by Maksimov et al. (2012). Formerly proposed binary peptide fragments are marked with colored asterisks.

Additional information on the thermal stability of the astexin-1 variants, including MS and MS2 spectra, is presented in Figure S2.

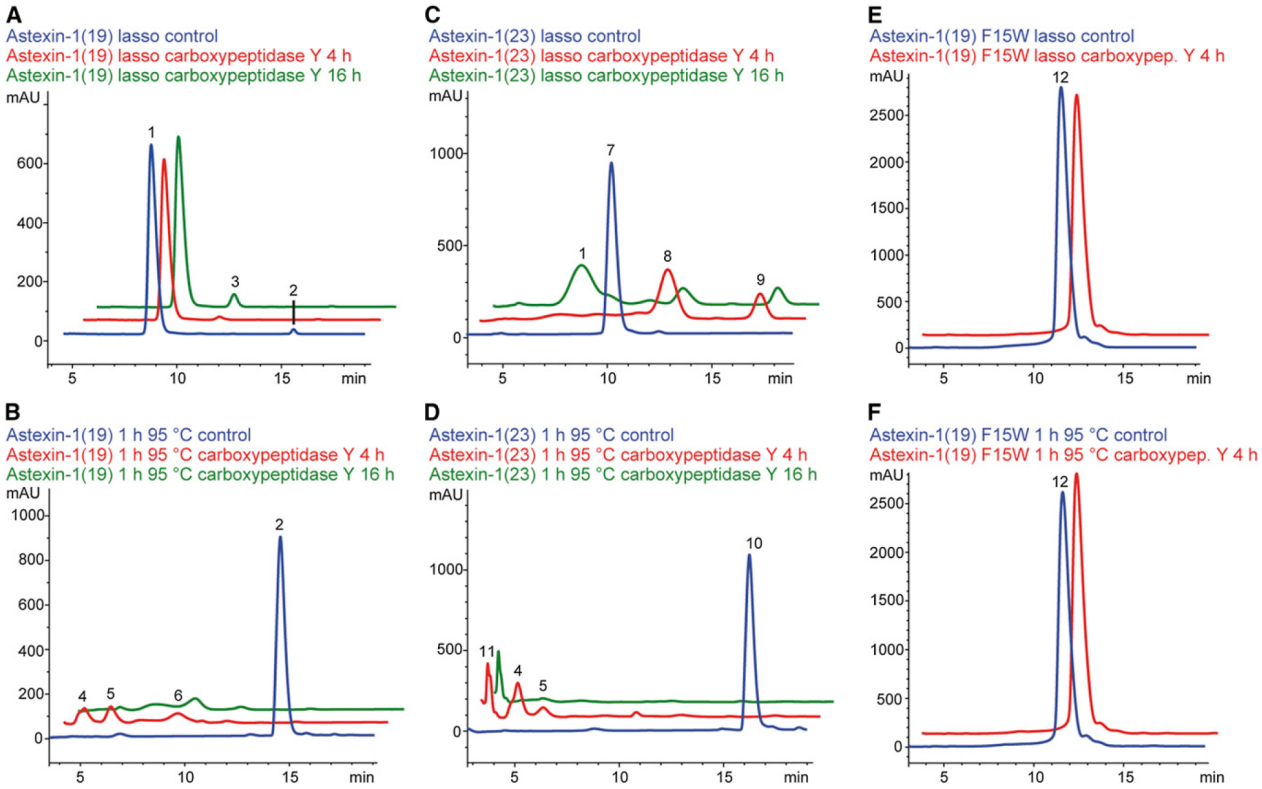


Figure 3. Carboxypeptidase Y Treatment of Native Astexin-1(19) and (23) and the Astexin-1(19) F15W Variant, Shown as HPLC Chromatograms

(A and B) Astexin-1(19) (A) before and (B) after heat treatment. **1** astexin-1(19) lasso, **2** astexin-1(19) unfolded, **3** astexin-1(18) lasso, **4** astexin-1(12) and (13), **5** astexin-1(11), and **6** astexin-1(10).

(C and D) Astexin-1(23) (C) before and (D) after heat treatment. **1** astexin-1(19) lasso, **4** astexin-1(12) and (13), **5** astexin-1(11), **7** astexin-1(23) lasso, **8** astexin-1(21) lasso, **9** astexin-1(20) lasso, **10** astexin-1(23) unfolded, and **11** unidentified.

(E and F) Astexin-1(19) F15W variant before (E) and after (F) heat treatment. **12** astexin-1(19) F15W lasso.

Additional information about astexin-1 stability against trypsin and of the variants stability against carboxypeptidase Y is presented in Figure S3.

becomes prone to proteolytic degradation. The treatment with carboxypeptidase is therefore a useful tool to distinguish between folded and unfolded state.

Carboxypeptidase Y was only able to very slowly cleave off one amino acid of astexin-1(19) in its lasso fold; however, when heat treated, the enzyme rapidly degraded the unfolded molecule (Figures 3A and 3B). As expected, the heat-unfolded full-length astexin-1(23) was found to be amenable to carboxypeptidase Y (Figure 3D), but surprisingly the untreated molecule was also found to be carboxypeptidase sensitive, although at a much slower rate, generating variants with 21, 20, and 19 residues (Figure 3C). Starting from astexin-1(23), no further degradation products shorter than astexin-1(19) were detected, indicating that the four C-terminal residues (INQD) of full-length astexin-1(23) are easily accessible to carboxypeptidase Y, whereas Arg19 seems to be more shielded but is also accessible according to the proteolytic behavior of astexin-1(19). This also shows why the Δ4 variant astexin-1(19) is one of the major fermentation products that build a stable lasso fold, although a shorter variant astexin-1(18) could be detected (Figure 1B).

To further access the role of Arg19, which is the last residue in astexin-1(19), trypsin digestion was performed. If Arg19, which is the only recognized residue for this protease in astexin-1, is truly the plug amino acid, it should be shielded by the ring and would not be accessible for degradation. Both lasso and unfolded astexin-1(23) were degraded by trypsin to the Δ4 variant astexin-1(19), with the unfolded variant being degraded slightly faster (Figures S3C and S3D available online). This finding is in accordance with the carboxypeptidase Y treatment and furthermore reveals the accessibility of the arginine residue at position 19, which was already suggested by the carboxypeptidase digestion of astexin-1(19).

Mutational Analysis of the Precursor Peptide

Several mutational studies of microcin J25, capistruin, and the recently discovered caulosegnins have shown that these are useful approaches to gain insights into the specificity of the biosynthetic machinery or to gain information about the mode of action of a lasso peptide (Cheung et al., 2010; Hegemann et al., 2013; Knappe et al., 2009; Pan and Link 2011; Pan et al., 2012b; Pavlova et al., 2008; Semenova et al., 2005). Therefore,

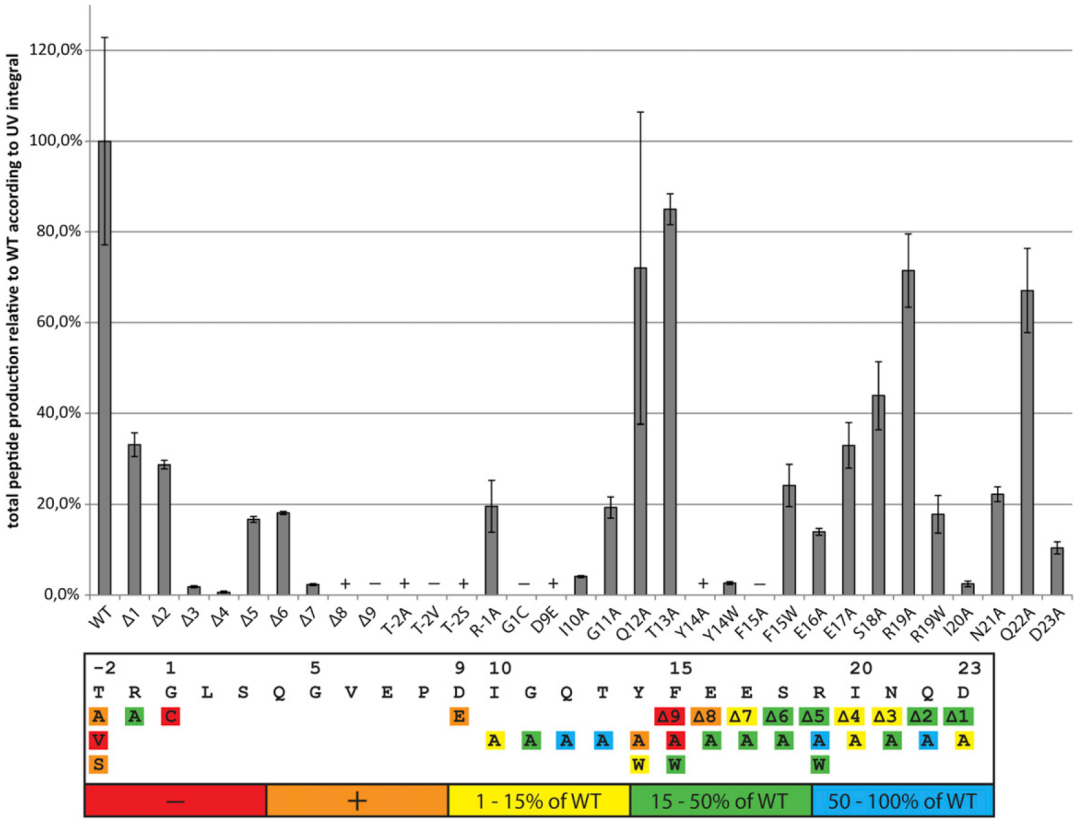


Figure 4. Production of Astexin-1 Variants Compared to Native Astexin-1 Production Level, Shown in a Color-Coded Schematic Representation

Values are average of three independent measurements; error bars show standard deviation. +, detected by MS; −, not detected by MS. A summary of the results of the mutational analysis, including the stability of all variants, is presented in Table S1.
See also Table S6.

several sets of mutants of astexin-1 were generated to study the specificity of the processing enzymes and to identify the location of important residues in the tail region. Finally, rational mutagenesis was applied to improve the thermal stability of this lasso peptide.

To gain knowledge about the tolerance of the maturation machinery of astexin-1, variants of the threonine, which is conserved in all known lasso peptide precursors, and the arginine in front of the cleavage site (T-2A, T-2V, T-2S, R-1A) were constructed. Furthermore the amino acids, which are involved in the macrolactam ring formation, were altered yielding variants G1C and D9E (Figure 4). The results obtained for the mutagenesis of the conserved Thr-2 were in contrast to the previously published observations for microcin J25, capistruin, and the caulosegnins (Hegemann et al., 2013; Pan et al., 2012b). From the three variants analyzed in this study, two were only detected in trace amounts (T-2A and T-2S), whereas the third exchange (T-2V) completely abolished astexin-1 production (Figure 4). This indicates that the astexin-1 maturation machinery is less tolerant for substitutions of this conserved threonine than the processing enzymes of capistruin, microcinJ25, and the caulosegnins. Although substitution of the Arg-1 (R-1A) had a less serious impact on peptide production, the detected amount of this variant was only 20% compared to the native system (Figure 4). Although this is in accordance with the qualitative mutational study of capistruin (Knappe et al., 2009), in which the His-1 could be exchanged for alanine, it indicates that this location seems to be not essential but influential for efficient recognition by the processing enzymes. Additionally, there is some general conservation of a positively charged residue (His, Lys, Arg) at this position in putative precursor peptides in proteobacteria, recently identified by genome mining (Maksimov et al., 2012; Severinov et al., 2007).

Mutation of the macrolactam ring-forming amino acids (Gly1 and Asp9) for microcin J25 was addressed by Pavlova et al. (2008) and for capistruin by Knappe et al. (2009), which in both cases revealed no tolerance for substitutions. For astexin-1 a similar result is presented here for Gly1, where an exchange to cysteine abolished production; however, a substitution of Asp9 by glutamic acid (D9E) generates a variant detectable by MS. The retention time, which should not significantly be altered by such a conserved mutation (Asp to Glu), was 3 min higher compared to the native peptide, and upon heat treatment of the extract, no change in retention time was observed

(comparable data shown in Figure S2B). Additionally, the proteolytic stability was elucidated by treating the dried pellet extract of a fermentation of this variant with carboxypeptidase Y. The following HPLC-MS analysis revealed the instability of the D9E variant, indicating that it very likely does not adopt a lasso fold (Figure S3K), which could also explain the detected low amounts. In context of the low thermal stability of native astexin-1, it is not surprising that through the expansion of the macrolactam ring size by one carbon atom (Asp9 to Glu9) in this variant, the lasso fold cannot be maintained anymore.

To elucidate the tolerance of substitutions for potential scaffold applications as well as the identification of important residues for the biosynthesis, single-residue alanine substitutions were constructed for all of the 14 residues of the extracyclic part of astexin-1 (Figure 4; Table S1). The overall tolerance against these substitutions was rather high, with exception of the Y14A and F15A variants, which could hardly or not at all be detected by HPLC-MS, respectively. Interestingly, most of the mutations altering the chemical and/or sterical nature of the amino acid (such as G11A, Q12A, T13A, E16A, E17A, S18A, R19A, N21A, Q22A, and D23A) had only little to moderate effects on astexin-1 production (Figure 4; Table S1).

In order to understand the importance of the residues Tyr14 and Phe15 and to explain the surprisingly well produced variant R19A (Figure 4; Table S1), combined tandem mass spectrometric analysis, thermal stability investigation, and carboxypeptidase digestion of several mutants were performed. According to its thermal and proteolytic behavior, the Y14A variant was found to be most likely already unfolded (Figures S2B and S3L), whereas the R19A variant adopted the lasso fold and was as heat labile as native astexin-1 (comparable results shown in Figure S2A), which strongly contradicts its formerly proposed role as the plug amino acid. To investigate the importance of the long tail and to identify potentially important residues, truncation mutants were generated and analyzed revealing that truncated variants from astexin-1(22) (Δ1) to astexin-1(15) (Δ8) could be detected by MS (Figure 4). Furthermore, the variants astexin-1(22) (Δ1) to astexin-1(16) (Δ7) could be investigated for thermal stability and astexin-1(17) to astexin-1(16) also for proteolytic stability with HPLC-MS and tandem mass spectrometry techniques. These studies revealed that all truncation variants (Δ1 to Δ7) are heat-labile lasso peptides as native astexin-1 (Figures S2A and S3E–S3H). Together these observations, in particular the low tolerance for substitution of Tyr14 and Phe15, the behavior of the Y14A variant, and the possibility of creating heavily truncated variants that behave almost like native astexin-1(23), strongly suggest that Tyr14 and Phe15 play a key role in stabilizing the lasso structure and additionally exclude Arg19 from being the only residue involved in maintenance of the lasso fold.

Because the native astexin-1 is quite heat labile, the question arose, whether it is possible to create a thermostable variant of this peptide. Therefore, three variants were generated by replacing the three largest residues in the tail (Tyr14, Phe15, and Arg19) with tryptophan. HPLC-MS analysis, tandem mass spectrometry, and carboxypeptidase Y digestion before and after heat treatment clearly show that F15W is a heat-stable variant of astexin-1 (Figures 3E and 3F; Figure S2C), whereas the other two variants Y14W and R19W behave like native astexin-1 under thermally denaturing conditions (Figures S3I and S3J). This confirms the assumption that Phe15 is the essential, bulky residue involved in the entrapment of the C-terminal tail and maintenance of the lasso fold.

Structure Elucidation Using NMR Spectroscopy

The gene cluster as well as an NMR 3D structure of astexin-1(23) was presented as a proof-of-principle example to the "sequence pattern matching" approach for the genome mining of lasso peptides developed by Maksimov et al. (2012). The structure (Protein Data Bank [PDB] ID code 2LTI) was determined in DMSO-d_6 at 295 K with the following spectroscopic and structural features: (1) astexin-1(23) possessed a lasso structure, (2) the isopeptide bond forming the macrolactam ring was between Gly1 and Asp9, (3) an eight-residue loop above and a six-residue tail under the macrolactam ring was defined, with Glu17 above, Ser18 passing, and Arg19 under the ring, (4) abnormal hydrogen bond interactions between the loop (above the ring) and the tail (under the ring) were identified (in Figure 6C of their article) to stabilize the lasso structure, and (5) a very small ^{1}H chemical shift dispersion for the amide protons (0.7 ppm) and the αH protons (1.0 ppm). The results of our rather extended mutational and stability studies on astexin-1(23) and astexin-1(19) are in accordance with features (1) and (2) but in discrepancy with feature (3). Features (4) and (5) are uncommon to most of the known NMR structures of lasso peptides (Xie and Marahiel, 2012).

These conflicting results and the fact that both astexin-1(19) and astexin-1(23) are highly hydrophilic prompted us to investigate the 3D structures of these peptides in aqueous solution thoroughly by using one- (1D) and two-dimensional (2D) NMR spectra. Thus, NMR spectra were recorded on sample **1**: 3.5 mg of astexin-1(23) dissolved in 250 μl of H_2O/D_2O (9:1) and sample **2**: 3.0 mg of astexin-1(19) dissolved in 250 μl of H_2O/D_2O (9:1). The samples were prepared in standard way and filled into 3 mm NMR tubes (see the Experimental Procedures for NMR experimental details). The concentrations of the samples were 5.5 and 5.7 mM, respectively. The ^{1}H spectra of sample **1** at temperatures between 278 and 298 K with 5 K increments are shown in Figure 5. Spectra are presented in the region 9.65–6.75 ppm for clarity, and labels for the signal assignments of the amide protons are attached. Neat and well-resolved single set of signals were observed, as reflected in Figure 5. A diverse distribution of the temperature response of these signals revealed a stable lasso folding of this peptide (Xie and Marahiel, 2012). The best signal resolution was observed at 288 K, and a full signal assignment (Table S2) was thus fulfilled by standard procedure (see below) at this temperature.

In a similar way, variable temperature ^{1}H spectra of sample **2** at the same temperatures were recorded (see Figure S4 for spectra). Close inspection on each pair of spectra at the same temperature revealed that the spectra of sample **2** appear to be those of sample **1** subtracted with signals of the last four residues of sample **1**, i.e., I20N21Q22D23. On the basis of the close similarity in variable temperature ^{1}H spectra of samples **1** and **2**, we concluded that the two peptides under the same conditions possess 3D structures with close similarity. Thus, the last four residues in the C-terminal tail of astexin-1(23) have little contribution to the core of its 3D structure under the applied experimental condition. Therefore, detailed structural determination was performed on astexin-1(19).

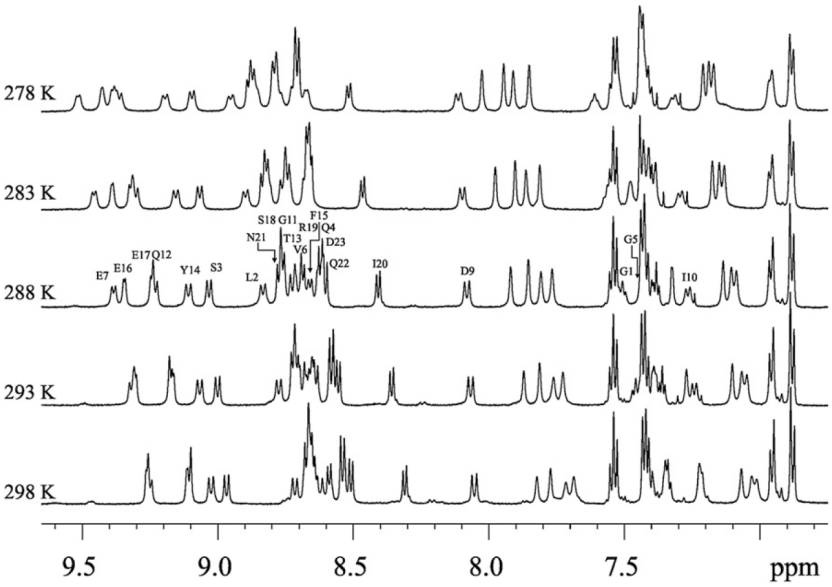


Figure 5. Variable Temperature ^{1}H Spectra in the Region 9.65–6.75 ppm of Astexin-1(23) in H_2O/D_2O at a Ratio of 9:1

The labels are the signal assignments of the amide protons. For comparison, the ^{1}H spectra at variable temperatures for astexin-1(19) are shown in Figure S4.

See also Tables S2 and S3.

All assignments of the ^{1}H signals were obtained by standard procedures (Wüthrich, 1986). A combination of DQF-COSY and nuclear Overhauser effect spectroscopy (NOESY) produced sequential assignments (i.e., all αH and NH and their sequence in the backbone), and a combination of DQF-COSY and TOCSY allowed the determination of the side chains. One pure conformation was observed. Full assignment of ^{1}H signals was thus obtained (NOESY spectrum: Figure S5; ^{1}H chemical shifts: Table S3). Strong NOE contacts between the NH of Gly1 and the βH of Asp9, showing an internal linkage between the two residues, were observed. Strong sequential αH–δH NOESY cross-peaks were detected, which showed the transconformation of Pro8. Inspection of ^{1}H spectra between 278 and 298 K (Figure S4) revealed almost no temperature dependence of the NH Asp9 and very weak dependence of those of Tyr14 and Phe15. Variable delay ^{1}H spectra in D_2O (data not shown) are consistent with a very slow deuterium exchange of the amide protons of these residues. Furthermore, a large number of long-range NOE contacts were observed. These are the $d_{NN}(i,j)$ connectivity between Ser3–Tyr14 and Glu7–Phe15, the $d_{\alpha N}(i,j)$ connectivity between Pro8–Phe15 and Tyr14–Asp9, the $d_{\beta N}(i,j)$ connectivity between Asp9–Gly1, Tyr14–Gly6, and Tyr14–Val6, and the $d_{\alpha\alpha}(i,j)$ connectivity between Gly1–Phe15 and Pro8–Tyr14. All these short distances revealed in the NOESY spectrum are in favor of the lasso structure of astexin-1(19).

Structure calculations were performed with the program CYANA (v. 2.1) (Herrmann et al., 2002). The internal linkage was realized by setting the distance constraints between N of Gly1 and Cγ of Asp9 to be 1.33 Å. NOE cross-peaks observed in the of 100 ms mixing time NOESY experiment were converted into distance constraints manually. In this way, 169 unambiguous distance constraints were obtained, 56 for the backbone, 24 for long-range interactions, and 89 for the side chains. Thus, there was an average of 8.9 distance constraints per residue. In addition, constraints of torsion angles ϕ and χ^1 were determined by analyzing the vicinal coupling constants $^{3}J_{HN\alpha}$ and $^{3}J_{\alpha\beta}$.

The above-mentioned constraints were used in the simulated annealing protocol for calculation in the CYANA (v. 2.1) program. The calculation initiated with 50 random conformers, and the resultant structures were engineered by the program package Sybyl (v. 7.3) (Tripos, 2006) to include the covalent linkage between the nitrogen of Gly1 and Cγ of Asp9, followed by energy minimization under NMR constraints using a TRIPOS force field within Sybyl. Thus, on the basis of low energies and minimal violations of the experimental data, a family of 20 structures was chosen. These 20 energy-minimized conformers show an average root-mean-square deviation (rmsd) of 0.03 Å and are kept to represent the solution structure of astexin-1(19) (Figure 6; PDB ID code 2M37).

The family of 20 structures shown in Figure 6 represents the lasso fold of astexin-1(19) in aqueous solution at 10°C. On the basis of the close similarity in the ^{1}H variable temperature spectra between astexin-1(19) and astexin-1(23), the lasso structure of 19 amino acids is assumed to represent the core of the 3D structure of astexin-1(23) in solution under the same conditions. Thus, our lasso structure shares the feature of the lasso ring being a nine-membered ring with that of astexin-1(23) defined by Maksimov et al. (2012). Furthermore, our structure shows the following properties, which are in discrepancy with the structure defined by Maksimov et al. (2012): (3) a five-membered loop above and a five-membered tail (for astexin-1(19)) below the macrolactam ring, whereby Tyr14 and Phe15 are located on opposite sides of the ring and serve as steric lock; (4) a short fragment of relative weak β sheet could be defined between E7P8 and Y14F15 on the basis of the observed long-range NOE contacts (Figure 6C); and (5) no regular turn could be assigned to the loop. All these structural features are in accordance with the properties identified in the mutagenesis and stability studies. Therefore, similar to microcin J25 (Bayro et al., 2003; Rosengren et al., 2003; Wilson et al., 2003), the steric lock by bulky residues Tyr14 and Phe15 and the formation of a β sheet are assumed to be the major factors for the stability of the lasso fold in astexin-1. One consistent lasso structure was identified for microcin J25 both in aqueous solution (Bayro et al., 2003; Rosengren et al., 2003) and in DMSO-d6 (Wilson et al., 2003). Thus, we concluded that our 3D structure of astexin-1(19) in aqueous solution represents the correct lasso folding of the title peptides of astexin-1 and raised the question of whether the original NMR data of astexin-1(23) in DMSO-d6

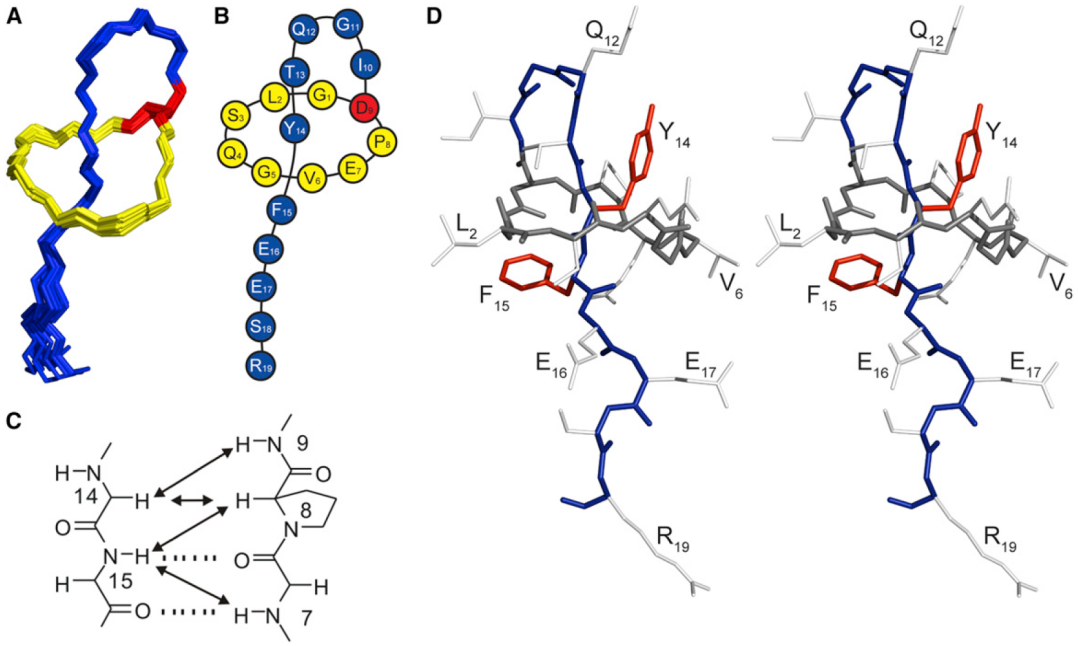


Figure 6. 3D NMR Structure of Astexin-1(19)

(A) Superimposition of the 20 lowest energy structures of astexin-1(19). Structures are superimposed over all backbone atoms. The isopeptide bond forming Asp9 is colored in red; the remaining ring amino acids are colored in yellow; and the tail residues are colored in blue.

(B) Schematic representation of the astexin-1(19) structure.

(C) Scheme of the small β sheet between E7-P8 and Y14-F15.

(D) Representative average structure of astexin-1(19) shown as relaxed eyes stereo picture. Backbone of the tail is colored in blue, and backbone of the ring is colored in dark gray. Side chains of the upper (Tyr14) and lower (Phe15) plug residues are shown in red, and other side chains are shown in light gray. The structure is tilted and turned for better visibility of residues. NOESY NMR spectra used for determination of the structure are shown in Figure S5.

See also Figure S7 and Table S4.

(Maksimov et al., 2012) were recorded on a natural peptide with the lasso fold. A quality comparison of the 20 minimum energy structures and their Ramachandran plots is shown in Figure S6.

Although the entrapment of the ring between the two residues Tyr14 and Phe15, shown in our NMR structure, appears to confer a rigid fixation of the tail motion, the heat stability investigation suggests otherwise. A mapping of the solvent-accessible surface on the average structure reveals a hole in the ring in a comparable size to the Phe15 side chain, suggesting a possible passage, which may explain the observed thermal sensitivity of the native lasso peptide as well as the thermal stability of the F15W variant (Figure 7). Even though only the structure of the truncated variant astexin-1(19) was elucidated, the similarity of the recorded NMR spectra of both variants was high enough to assume that full-length astexin-1(23) adopts the same fold. With a length of nine amino acids, the tail of astexin-1(23) is therefore the longest tail below the ring of all lasso peptides known so far.

In conclusion, we successfully obtained two main products from the recently identified biosynthetic gene cluster of the lasso peptide astexin-1. Both astexin-1(19) and astexin-1(23) were shown to be sensitive to thermal unfolding, while exhibiting certain stability against proteolytic degradation by carboxypeptidase Y and trypsin. Mutational analysis of the leader and core sequence of the precursor peptide provided important insights into the specificity of the biosynthetic machinery and revealed some hot spots that have great influence on in vivo production and in vitro stability, in the sequence of this lasso peptide. Finally, we performed a thorough NMR spectroscopic study on astexin-1(23) and astexin-1(19) and determined the 3D structure of astexin-1(19) in aqueous solution.

SIGNIFICANCE

Despite the predictability of their gene clusters by a genome-mining-centered approach, lasso peptides are still a young class of ribosomally assembled natural products, with only few known representatives. In this study, we investigate the structure-stability relationship of the lasso peptide astexin-1, revealing it to be one of the few examples in its class to be a heat-sensitive molecule. Further investigation by mutational analysis showed a broad tolerance of the astexin-1 biosynthetic machinery for most residues located in the tail region, with some exceptions of hot spots, whereas mutations at residues involved in the macrolactam ring formation or around the protease cut site within the leader sequence drastically influenced the lasso peptide production in vivo. Astexin-1 is an example of a lasso peptide for which it was possible to exchange the macrolactam ring-forming acidic residue yielding a branched cyclic

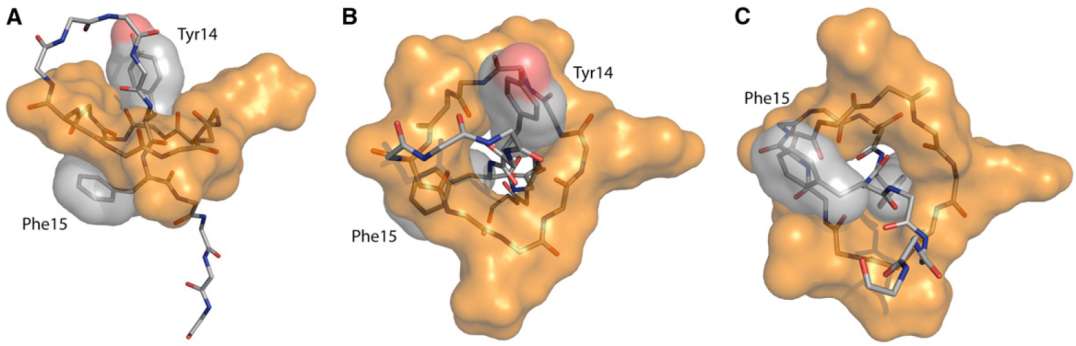


Figure 7. Interactions of the C-Terminal Tail and the Macrocyclic Ring of Astexin-1(19)

(A–C) 3D structure of astexin-1(19), where entire backbone and side chains of the plug amino acids Y14 and F15 are shown as sticks, and complete macrolactam ring and plug side chains (Y14, F15 on opposite sides of the macrolactam ring) are shown as solvent-accessible surfaces. View (A) from the side, (B) from above, and (C) from below the ring. Isopeptide bond in each panel is on the left-hand side. Quality comparison of structures and Ramachandran plots of the current and the recently published structure are shown in Figure S6.

peptide. Variants carrying mutations at identified hot spots were further studied regarding their in vitro stability. However, none of the observed behaviors concerning thermal stability of parental astexin-1 or that of the generated variants was compatible with the recently published structure of astexin-1 (Maksimov et al., 2012). This inconsistency encouraged us to reinvestigate the astexin-1 structure. Our NMR studies on astexin-1(19) revealed a tight structure with a short loop region above the nine-membered macrolactam ring and a long tail comprising of nine residues for the full-length peptide of which Phe15, and not Arg19 (as published by Maksimov et al. 2012), acts as the plug residue to trap the tail. Understanding these molecular aspects of the structure-stability interactions of astexin-1 allowed the conversion of this heat-labile natural product into a thermostable one by rational design.

EXPERIMENTAL PROCEDURES

Strains and general methods, such as cloning, construction of mutants, and heterologous expression, are described in the Supplemental Experimental Procedures.

Purification of Astexin-1 and Variants

After centrifugation (6,000 rpm), the cell pellet was combined and extracted with methanol for 1 hr at room temperature (RT). The solvent obtained after extraction was evaporated to dryness with a rotational evaporator at 30°C, or in case of the variants, it was removed via lyophillization.

Crude extracts of native astexin-1 were resuspended in 20% acetonitrile and subsequently applied to a preparative RP-HPLC system (1100 series, Agilent, Santa Clara, CA, USA) using a C18HTec column (250 × 21 mm) with the following gradient of water/0.1% trifluoroacetic acid (solvent A) and acetonitrile/0.1% trifluoroacetic acid (solvent B) with a flow rate of 18 ml/min: a linear increase from 20% B to 30% B in 30 min followed by the washing increasing to 95% B in 2 min and holding 95% B for 5 min. Retention times (Rt) of the produced compounds were between 8 and 19 min (Figure S1A). Fractions containing astexin-1(23) and (19) were separately combined, lyophilized, and subjected to another HPLC purification step.

Further purification of the products was performed with an analytical scale HPLC system (1260 series, Agilent) with a fraction collector using a Nucleodur C18ec column (125 × 2 mm) at a column temperature of 50°C and a flow rate of 0.2 ml/min with the following gradient using the same solvents as before: linear increase from 20% B to 27.5% B in 15 min followed by the washing performed in the same manner as the preparative scale. The truncated product astexin-1(19) had an Rt of 8.2 min, and the full-length product astexin-1(23) had an Rt of 10.3 min (Figure S1B). Fractions were combined and lyophilized, and in case of astexin-1(23), they were resuspended in 15% acetonitrile and subjected to a third HPLC purification step.

Final purification of the full-length product was performed with the analytical scale HPLC system using a polar RP column (Synergi [Phenomenex, Torrance, CA, USA], 4μ 250 × 2 mm) at a column temperature of 25°C and a flow rate of 0.5 ml/min with the following gradient using the same solvents as before: 17% B is held for 15 min and then washing is applied in the same manner as before. The retention time for the full-length product was 6.4 min (Figure S1C). The final yield of astexin-1(23) was approximately 4 mg/l culture, and that of astexin-1(19) was about 5 mg/l.

Variants were purified with the preparative HPLC system. For variants Δ6 aa and F15W, the same gradient as for the native astexin-1 was used. Variants Δ7 aa and R19W were purified using a gradient going from 25% to 35% solvent B in 30 min, whereas all other steps were performed as mentioned before.

Mass Spectrometric Analysis

Mass spectrometric analysis of astexin-1 and variant extracts were performed with a LTQ-FT ultrainstrument (Thermo Fisher Scientific, Waltham, MA, USA) connected to a micropore 1100 HPLC system (Agilent) using sample amounts of up to 100 μl. For detection, the UV absorption at 215 nm was recorded. Separation was achieved using a CC Nucleosil 300-8 C18 column (125 × 2 mm) (Macherey-Nagel, Dueren, Germany), applying the following gradient of water/0.1% trifluoroacetic acid (solvent A) and acetonitrile/0.1% trifluoroacetic acid (solvent B) at a column temperature of 40°C and a flow rate of 0.2 ml/min: starting with a linear increase from 5% to 50% B in 30 min, a subsequent increase from 50% to 95% B in 2 min, and holding 95% B for additional 5 min.

For product quantification, UV peak areas were integrated and relative production was determined by comparison to the UV peak area of the corresponding wild-type.

Collision-induced dissociation fragmentation studies within the linear ion trap were done using online HPLC-MS. In most cases, the doubly charged ions were selected for fragmentation, as they were the dominant species in the spectra, while for shorter variants also the singly charged species could be analyzed. The energy for fragmentation was set to 35 for every measurement performed.

Mass spectrometric analysis after the thermal stability and protease assays of the purified lasso peptides and variants thereof was performed with a 1100 series MSD (Hewlett-Packard) coupled with a micropore 1260 HPLC system (Agilent).

For astexin-1(19) and astexin-1(23) and variants Δ6 and F15W, the HPLC-MS runs were performed with the same gradient as in the second purification step of astexin-1. An adapted gradient was used for variant Δ7, going from 22.5% B to 30% B in 15 min. For variant R19W, the gradient was further adjusted, starting from 25% B and finishing with 32.5% B. The washing steps were applied analog to the purification.

MALDI-TOF analysis was performed with a 4800 MALDI TOF/TOF analyzer (AB SCIEX, Framingham, MA, USA). Samples of folded and thermally unfolded astexin-1(23) were mixed with dihydroxybenzoic acid as the matrix on a MALDI target, dried for 15 min, and applied to the spectrometer.

Heat Stability Assays of Astexin-1 and Variants

To investigate the thermal stability of the main lasso peptide products astexin-1(23) and astexin-1(19), a solution of 10 μg purified lasso peptide was incubated at 95°C for either 1, 2, 4, or 8 hr or for 4 hr at either 20°Ca, 35°C, 50°C, 65°C, 80°C, or 95°C. Samples were cooled to 4°C and were subsequently analyzed via HPLC-MS using the same gradient as in the second purification step of astexin-1.

Heat stability of the variants was investigated by incubating 25 μl of the respective extract at 95°C for 1 hr. The samples were cooled and analyzed via HPLC-MS. As reference, 25 μl of the respective untreated extract were analyzed.

Protease Assays of Astexin-1 and Variants

Stability against proteolytic degradation of the lasso peptides and the branched cyclic variants were investigated by incubating 10 μg of the purified lasso peptide before and after heat treatment (95°C, 1 hr) with different proteases.

The carboxypeptidase Y assays were performed with 0.5 U carboxypeptidase Y in a buffer containing 50 mM MES and 1 mM $CaCl_2$ at a pH of 6.75 for 4 hr or 16 hr at 25°C. For variants that were not isolated because of low production, 25 μl of the pellet extract were evaporated to dryness at a concentrator (model 5301, Eppendorf, Hamberg, Germany), resuspended in a carboxypeptidase Y (1.0 U) containing buffer (50 mM MES and 1 mM $CaCl_2$ [pH 6.75]), and incubated for 2 hr at 25°C. The trypsin assays were performed with 0.5 μg trypsin in a buffer containing 50 mM Tris-HCl and 1 mM $CaCl_2$ at a pH of 7.6 for 4 hr at 25°C.

Distinction between Lasso Peptide and Branched-Cyclic Peptide

If not further stated in the paper, the two topologies of a peptide were distinguished by the difference in their retention time and their fragmentation patterns before and after heat treatment as well as their stability against digestion by carboxypeptidase Y. Heat treatment of the two lasso peptide main products and almost all variants generated showed a second peak after HPLC separation with a significantly higher retention time. This second peak, in contrast to the first one, is sensitive to carboxypeptidase Y digestion and can therefore be classified as a branched cyclic peptide, whereas the first compound, which has the lasso fold, could only be digested to a certain extent or not at all depending on the length of the tail. Furthermore, both topologies have distinguishable fragmentation patterns in tandem mass spectrometry. Although no binary peptide fragments have been observed, the general fragmentation of a lasso peptide is weaker than that of a branched cyclic peptide. In general, the lasso peptides show larger fragments, and the most intensive peak corresponds to the loss of neutral molecules like H_2O and NH_3.

NMR Spectroscopy

Three samples were used for the NMR spectroscopic studies. Sample **1** contained 3.5 mg of astexin-1(23) in 250 μl H_2O/D_2O (9:1), and sample **2** contained 3.0 mg of astexin-1(19) in 250 μl H_2O/D_2O (9:1), whereas sample **3** was 3.0 mg of astexin-1(19) in 250 μl D_2O. Samples were filled in Wilmad 3 mm tubes (Rototec Spintec, Griesheim, Germany). All the peptides studied were at natural abundance. Spectra were recorded on a Bruker Avance 600 MHz spectrometer equipped with an inverse triple-resonance ^{1}H-^{13}C-^{15}N probe with z-gradient. Temperature influence on samples **1** and **2** was studied by recording ^{1}H spectra at temperatures between 278 and 298 K with a 5 K increment. The 2D experiments DQF-COSY (Rance et al., 1983), TOCSY (Bax and Davis, 1985), and NOESY (Jeener et al., 1979) were performed in the standard way, with water suppression using excitation sculpting method (Hwang and Shaka, 1995) on samples **1** and **2** for sequential assignments. For sample **1**, the sequential assignment was done at 288 K. Because of crowdedness and partial overlap of the signals, further 2D spectra recorded at 283 K were used to assist the assignment. Sample **2** has a smaller size, thus its sequential assignment was easily fulfilled at 283 K. The TOCSY spectra were recorded at an 80 ms mixing time, and NOESY spectra were recorded at 100 and 300 ms mixing times. 1D spectra were acquired with 65,536 data points, whereas 2D spectra were collected using 4,096 points in the F_2 dimension and 512 increments in the F_1 dimension. Both 1D and 2D spectra were recorded with 32 transients and a relaxation delay of 3.0 s. Chemical shifts were referenced to H_2O signal, which in turn was calibrated using 2,2-dimethyl-2-silapentane-5-sulfonate (DSS) as internal standard in a different sample at the same temperature. To study the proton/deuterium exchange, ^{1}H and TOCSY spectra were recorded on sample **3** at 283 K sequentially 1 hr, 1 day, 7 days, and 28 days after sample preparation. All spectra were processed with Bruker TOPSPIN (v. 3.1). NOE cross-peaks were analyzed within the program Sparky (Goddard and Kneller, 2007).

Structure Calculations

Structure calculation was performed on astexin-1(19) in H_2O/D_2O (9:1) at 283 K. NOESY cross-peaks with mixing time 100 ms were used to create distance constraints. Constraints of torsion angle ϕ were determined by analyzing the vicinal coupling constants $^{3}J_{HN\alpha}$, whereas those of χ^{1} were determined by fulfilling stereospecific assignment on the basis of NOE contacts and $^{3}J_{\alpha\beta}$ (Wagner, 1990). These constraints were used in the simulated annealing protocol, and the structure calculations were done with the program CYANA 2.1 (Herrmann et al., 2002). The coordinates of the 20 lowest energy structures to present the solution structure of astexin-1(19) have been submitted to the Protein Data Bank (PDB) and assigned the accession number 2M37.

ACCESSION NUMBERS

The Protein Data Bank accession number for the coordinates of the 20 lowest energy structures to present the solution structure of astexin-1(19) is 2M37.

SUPPLEMENTAL INFORMATION

Supplemental Information includes seven figures, six tables, and Supplemental Experimental Procedures and can be found with this article online at http://dx.doi.org/10.1016/j.chembiol.2013.03.013.

ACKNOWLEDGMENTS

We would like to acknowledge Dr. Uwe Linne and all members of the mass spectrometry department of the Philipps-University Marburg for all performed measurements and fruitful discussions. Financial support from the Deutsche Forschungsgemeinschaft and the LOEWE program of the State of Hessen are gratefully acknowledged.

Received: January 14, 2013
Revised: March 8, 2013
Accepted: March 22, 2013
Published: April 18, 2013

REFERENCES

Adelman, K., Yuzenkova, J., La Porta, A., Zenkin, N., Lee, J., Lis, J.T., Borukhov, S., Wang, M.D., and Severinov, K. (2004). Molecular mechanism of transcription inhibition by peptide antibiotic Microcin J25. Mol. Cell *14*, 753–762.

Arnison, P.G., Bibb, M.J., Bierbaum, G., Bowers, A.A., Bugni, T.S., Bulaj, G., Camarero, J.A., Campopiano, D.J., Challis, G.L., Clardy, J., et al. (2013). Ribosomally synthesized and post-translationally modified peptide natural products: overview and recommendations for a universal nomenclature. Nat. Prod. Rep. *30*, 108–160.

Bax, A., and Davis, D.G. (1985). Mlev-17-based two-dimensional homonuclear magnetization transfer spectroscopy. J. Magn. Reson. *65*, 355–360.

Bayro, M.J., Mukhopadhyay, J., Swapna, G.V., Huang, J.Y., Ma, L.C., Sineva, E., Dawson, P.E., Montelione, G.T., and Ebright, R.H. (2003). Structure of antibacterial peptide microcin J25: a 21-residue lariat protoknot. J. Am. Chem. Soc. *125*, 12382–12383.

Cheung, W.L., Pan, S.J., and Link, A.J. (2010). Much of the microcin J25 leader peptide is dispensable. J. Am. Chem. Soc. *132*, 2514–2515.

Duquesne, S., Destoumieux-Garzón, D., Peduzzi, J., and Rebuffat, S. (2007). Microcins, gene-encoded antibacterial peptides from enterobacteria. Nat. Prod. Rep. *24*, 708–734.

Fréchet, D., Guitton, J.D., Herman, F., Faucher, D., Helynck, G., Monegier du Sorbier, B., Ridoux, J.P., James-Surcouf, E., and Vuilhorgne, M. (1994). Solution structure of RP 71955, a new 21 amino acid tricyclic peptide active against HIV-1 virus. Biochemistry *33*, 42–50.

Goddard, T.D., and Kneller, D.J. (2007). Sparky 3. Computer Graphics Laboratory or Resource for Biocomputing, Visualization, and Informatics (San Fransisco: University of California).

Hegemann, J.D., Zimmermann, M., Xie, X., and Marahiel, M.A. (2013). Caulosegnins I-III: a highly diverse group of lasso peptides derived from a single biosynthetic gene cluster. J. Am. Chem. Soc. *135*, 210–222.

Herrmann, T., Güntert, P., and Wüthrich, K. (2002). Protein NMR structure determination with automated NOE assignment using the new software CANDID and the torsion angle dynamics algorithm DYANA. J. Mol. Biol. *319*, 209–227.

Hwang, T.L., and Shaka, A.J. (1995). Water suppression that works - excitation sculpting using arbitrary wave-forms and pulsed-field gradients. J. Magn. Reson. A *112*, 275–279.

Inokoshi, J., Matsuhama, M., Miyake, M., Ikeda, H., and Tomoda, H. (2012). Molecular cloning of the gene cluster for lariatin biosynthesis of Rhodococcus jostii K01-B0171. Appl. Microbiol. Biotechnol. *95*, 451–460.

Iwatsuki, M., Tomoda, H., Uchida, R., Gouda, H., Hirono, S., and Omura, S. (2006). Lariatins, antimycobacterial peptides produced by Rhodococcus sp. K01-B0171, have a lasso structure. J. Am. Chem. Soc. *128*, 7486–7491.

Iwatsuki, M., Uchida, R., Takakusagi, Y., Matsumoto, A., Jiang, C.L., Takahashi, Y., Arai, M., Kobayashi, S., Matsumoto, M., Inokoshi, J., et al. (2007). Lariatins, novel anti-mycobacterial peptides with a lasso structure, produced by Rhodococcus jostii K01-B0171. J. Antibiot. *60*, 357–363.

Jeener, J., Meier, B.H., Bachmann, P., and Ernst, R.R. (1979). Investigation of exchange processes by 2-dimensional NMR-spectroscopy. J. Chem. Phys. *71*, 4546–4553.

Katahira, R., Shibata, K., Yamasaki, M., Matsuda, Y., and Yoshida, M. (1995). Solution structure of endothelin B receptor selective antagonist RES-701-1 determined by 1H NMR spectroscopy. Bioorg. Med. Chem. *3*, 1273–1280.

Kimura, K., Kanou, F., Takahashi, H., Esumi, Y., Uramoto, M., and Yoshihama, M. (1997). Propeptin, a new inhibitor of prolyl endopeptidase produced by Microbispora. I. Fermentation, isolation and biological properties. J. Antibiot. *50*, 373–378.

Knappe, T.A., Linne, U., Robbel, L., and Marahiel, M.A. (2009). Insights into the biosynthesis and stability of the lasso peptide capistruin. Chem. Biol. *16*, 1290–1298.

Knappe, T.A., Linne, U., Xie, X., and Marahiel, M.A. (2010). The glucagon receptor antagonist BI-32169 constitutes a new class of lasso peptides. FEBS Lett. *584*, 785–789.

Kuznedelov, K., Semenova, E., Knappe, T.A., Mukhamedyarov, D., Srivastava, A., Chatterjee, S., Ebright, R.H., Marahiel, M.A., and Severinov, K. (2011). The antibacterial threaded-lasso peptide capistruin inhibits bacterial RNA polymerase. J. Mol. Biol. *412*, 842–848.

Lin, P.F., Samanta, H., Bechtold, C.M., Deminie, C.A., Patick, A.K., Alam, M., Riccardi, K., Rose, R.E., White, R.J., and Colonno, R.J. (1996). Characterization of siamycin I, a human immunodeficiency virus fusion inhibitor. Antimicrob. Agents Chemother. *40*, 133–138.

Maksimov, M.O., Pelczer, I., and Link, A.J. (2012). Precursor-centric genome-mining approach for lasso peptide discovery. Proc. Natl. Acad. Sci. USA *109*, 15223–15228.

Nar, H., Schmid, A., Puder, C., and Potterat, O. (2010). High-resolution crystal structure of a lasso Peptide. ChemMedChem *5*, 1689–1692.

Pan, S.J., and Link, A.J. (2011). Sequence diversity in the lasso peptide framework: discovery of functional microcin J25 variants with multiple amino acid substitutions. J. Am. Chem. Soc. *133*, 5016–5023.

Pan, S.J., Rajniak, J., Cheung, W.L., and Link, A.J. (2012a). Construction of a single polypeptide that matures and exports the lasso peptide microcin J25. ChemBioChem *13*, 367–370.

Pan, S.J., Rajniak, J., Maksimov, M.O., and Link, A.J. (2012b). The role of a conserved threonine residue in the leader peptide of lasso peptide precursors. Chem. Commun. (Camb.) *48*, 1880–1882.

Pavlova, O., Mukhopadhyay, J., Sineva, E., Ebright, R.H., and Severinov, K. (2008). Systematic structure-activity analysis of microcin J25. J. Biol. Chem. *283*, 25589–25595.

Potterat, O., Wagner, K., Gemmecker, G., Mack, J., Puder, C., Vettermann, R., and Streicher, R. (2004). BI-32169, a bicyclic 19-peptide with strong glucagon receptor antagonist activity from Streptomyces sp. J. Nat. Prod. *67*, 1528–1531.

Rance, M., Sørensen, O.W., Bodenhausen, G., Wagner, G., Ernst, R.R., and Wüthrich, K. (1983). Improved spectral resolution in cosy 1H NMR spectra of proteins via double quantum filtering. Biochem. Biophys. Res. Commun. *117*, 479–485.

Rosengren, K.J., Clark, R.J., Daly, N.L., Göransson, U., Jones, A., and Craik, D.J. (2003). Microcin J25 has a threaded sidechain-to-backbone ring structure and not a head-to-tail cyclized backbone. J. Am. Chem. Soc. *125*, 12464–12474.

Salomón, R.A., and Farías, R.N. (1992). Microcin 25, a novel antimicrobial peptide produced by Escherichia coli. J. Bacteriol. *174*, 7428–7435.

Semenova, E., Yuzenkova, Y., Peduzzi, J., Rebuffat, S., and Severinov, K. (2005). Structure-activity analysis of microcinJ25: distinct parts of the threaded lasso molecule are responsible for interaction with bacterial RNA polymerase. J. Bacteriol. *187*, 3859–3863.

Severinov, K., Semenova, E., Kazakov, A., Kazakov, T., and Gelfand, M.S. (2007). Low-molecular-weight post-translationally modified microcins. Mol. Microbiol. *65*, 1380–1394.

Tripos. (2006). Sybyl 7.3 computer program. Tripos, St. Louis. http://www.tripos.com/.

Wagner, G. (1990). NMR investigations of protein-structure. Prog. Nucl. Mag. Res. Sp. *22*, 101–139.

Weber, W., Fischli, W., Hochuli, E., Kupfer, E., and Weibel, E.K. (1991). Anantin—a peptide antagonist of the atrial natriuretic factor (ANF). I. Producing organism, fermentation, isolation and biological activity. J. Antibiot. *44*, 164–171.

Wilson, K.A., Kalkum, M., Ottesen, J., Yuzenkova, J., Chait, B.T., Landick, R., Muir, T., Severinov, K., and Darst, S.A. (2003). Structure of microcin J25, a peptide inhibitor of bacterial RNA polymerase, is a lassoed tail. J. Am. Chem. Soc. *125*, 12475–12483.

Wüthrich, K. (1986). NMR of Proteins and Nucleic Acids (New York: Wiley).

Wyss, D.F., Lahm, H.W., Manneberg, M., and Labhardt, A.M. (1991). Anantin—a peptide antagonist of the atrial natriuretic factor (ANF). II. Determination of the primary sequence by NMR on the basis of proton assignments. J. Antibiot. *44*, 172–180.

Xie, X., and Marahiel, M.A. (2012). NMR as an effective tool for the structure determination of lasso peptides. ChemBioChem *13*, 621–625.

Yan, K.P., Li, Y., Zirah, S., Goulard, C., Knappe, T.A., Marahiel, M.A., and Rebuffat, S. (2012). Dissecting the maturation steps of the lasso peptide microcin J25 in vitro. ChemBioChem *13*, 1046–1052.

Yano, K., Yamasaki, M., Yoshida, M., Matsuda, Y., and Yamaguchi, K. (1995). RES-701-2, a novel and selective endothelin type B receptor antagonist produced by Streptomyces sp. II. Determination of the primary structure. J. Antibiot. *48*, 1368–1370.

Yano, K., Toki, S., Nakanishi, S., Ochiai, K., Ando, K., Yoshida, M., Matsuda, Y., and Yamasaki, M. (1996). MS-271, a novel inhibitor of calmodulin-activated myosin light chain kinase from Streptomyces sp.—I. Isolation, structural determination and biological properties of MS-271. Bioorg. Med. Chem. *4*, 115–120.

3.3 Lasso Peptides From Proteobacteria: Genome Mining Employing Heterologous Expression and Mass Spectrometry

Julian D. Hegemann, Marcel Zimmermann, Shaozhou Zhu, Dennis Klug, Mohamed A. Marahiel,
Lasso Peptides From Proteobacteria: Genome Mining Employing Heterologous Expression and Mass Spectrometry,
Biopolymers Pept. Sci. **2013**, *100(5)*, 527-542.
doi: 10.1002/bip.22326

Author contributions:

The project was designed by Julian D. Hegemann and Mohamed A. Marahiel with the aid of Marcel Zimmermann. The experiments were performed predominantly by Julian D. Hegemann himself. Additional experimental assistance was received by Marcel Zimmermann, Shaouhou Zhu and Dennis Klug in regard to the cloning of some of the constructs as well as help during the heterologous expression of the lasso peptides. The data obtained this way was analyzed and interpreted by Julian D. Hegemann with feedback of Marcel Zimmermann. The manuscript was written by Julian D. Hegemann and its revision was done together with Mohamed A. Marahiel and Marcel Zimmermann. In addition to this, Mohamed A. Marahiel was further responsible for the project supervision.

Background and Summary:

After it had previously been shown that lasso peptide production could already be significantly increased by simple incorporation of an *E. coli* optimized ribosomal binding site in the intergenic region between the precursor peptide and the B protein encoding genes,[22,64] the applicability of this approach was assessed for ten other putative lasso peptide biosynthetic gene clusters. These clusters were identified by a B-protein centric genome mining approach and all of them were of proteobacterial origin. The heterologous production of all twelve predicted lasso peptides was successful, even though the overall yields strongly differed between the different systems. Besides showing that the incorporation of the ribosomal binding site was beneficial in most of the tested systems, this study was also concerned with providing fast and simple means, with which it would be possible to provide strong evidence for the presence of a lasso fold without having to rely on an NMR or X-ray based three-dimensional structure elucidation. Thus, a general procedure that used thermal stability assays in combination with protease treatment and high resolution MS-based analysis was developed, which allowed to quickly and efficiently gain strong evidence for the presence or absence of a lasso fold and at the same time was able to prove if the lasso peptide in question was heat sensitive. In light of the expansion of the number of known functional lasso peptide biosynthetic gene clusters from five to fifteen, the corresponding processing enzymes were also successfully analyzed for the presence of any conserved motifs, which can not only facilitate future genome mining approaches, but also substantiated previous studies about the catalytic activity of these kinds of enzymes.[30,59-61] Finally, a thorough bioinformatic analysis of the 98 identified putative lasso peptide biosynthetic gene clusters from proteobacteria and their surroundings allowed us to identify several distinct groups of cluster organizations. The largest of these groups where the ABCD arrangement known from the capistruin system and the plain ABC arrangement that is predominantly found adjacent to an inverse facing gene cluster comprising of four conserved genes encoding a σ-factor, an anti-σ-factor, a TonB-dependent receptor and a peptidase. The significance of the latter observation was later shown by a study done by Link and co-workers, who proved that the peptidase of such a gene cluster is indeed a lasso peptide specific isopeptidase.[23]

Lasso Peptides From Proteobacteria: Genome Mining Employing Heterologous Expression and Mass Spectrometry

Julian D. Hegemann, Marcel Zimmermann, Shaozhou Zhu, Dennis Klug, Mohamed A. Marahiel
Department of Chemistry, Philipps-University Marburg, Hans-Meerwein-Strasse 4 and LOEWE-Center for Synthetic Microbiology, Marburg D-35032, Germany

Received 14 January 2013; revised 31 May 2013; accepted 5 June 2013
Published online 29 July 2013 in Wiley Online Library (wileyonlinelibrary.com). DOI 10.1002/bip.22326

ABSTRACT:

Lasso peptides are natural products with a unique three dimensional structure resembling a lariat knot. They are from ribosomal origin and are post-translationally modified by two enzymes (B and C), one of which shares little similarity to enzymes outside of lasso peptide biosynthetic gene clusters and as such is a useful target for genome mining. In this study, we demonstrate a B protein–centric genome mining approach through which we were able to identify 102 putative lasso peptide biosynthetic gene clusters from a total of 87 different proteobacterial strains. Ten of these clusters were cloned into the pET41a expression vector, optimized through incorporation of a ribosomal binding site and heterologously expressed in Escherichia coli BL21(DE3). *All 12 predicted lasso peptides (namely burhizin, caulonodin I, caulonodin II, caulonodin III, rhodanodin, rubrivinodin, sphingonodin I, sphingonodin II, syanodin I, sphingopyxin I, sphingopyxin II, and zucinodin) were detected by high-resolution Fourier transform mass spectrometry and their proposed primary structure was confirmed through tandem mass spectrometry. High yields (ranging from 0.4 to 5.2 mg/L) were observable for eight of these compounds, while thermostability assays revealed five new representatives of heat labile lasso peptides.*

Keywords: genome mining; lasso peptide; thermal stability; mass spectrometry; proteobacteria

This article was originally published online as an accepted preprint. The "Published Online" date corresponds to the preprint version. You can request a copy of the preprint by emailing the Biopolymers editorial office at biopolymers@wiley.com

INTRODUCTION

Lasso peptides are an interesting group of natural products that share a unique structural motif. All known lasso peptides consist of an N-terminal macrolactam ring, which is eight to nine amino acids in size and is formed between the N-terminal amino group of a glycine or cysteine residue at position 1 and the side chain of a glutamate or aspartate residue at position 8 or 9.[1–16] Their distinguishing feature is the threading of the C-terminal tail through the macrolactam ring, yielding a structure reminiscent of a lariat knot.[5,12,13,15,17–26]

This entropically unfavorable conformation is sustained by the placement of sterically demanding residues (plug amino acids) above and below the ring, entrapping the tail inside and preventing its unthreading. Maintenance of the lasso fold by such kinetic means leads to a very confined structure and can confer tremendous stability against heat, chemical denaturation agents and proteases.[2,5,8,13,15,17,21,27–30] Nonetheless, the recently reported lasso peptides caulosegnin I and III[13] as well as a capistruin

Additional Supporting Information may be found in the online version of this article.

Correspondence to: Mohamed A. Marahiel, Department of Chemistry, Philipps-University Marburg, Hans-Meerwein-Strasse 4 and LOEWE-Center for Synthetic Microbiology, Marburg D-35032, Germany; e-mail: marahiel@staff.uni-marburg.de

Contract grant sponsors: Deutsche Forschungsgemeinschaft and the LOEWE Program of the State of Hesse

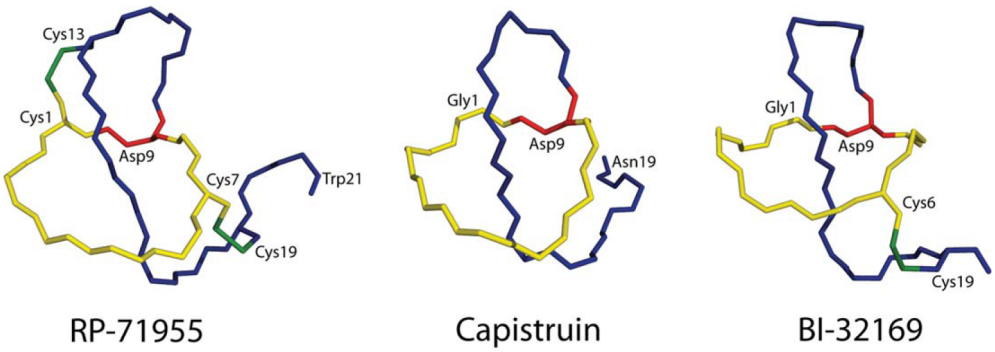


FIGURE 1 General classification criteria for lasso peptides. The representatives shown for each class are RP–71955[17] (class I, 4 cysteines, 2 disulfide bonds, Cys1), capistruin[12] (class II, 0 cysteines, 0 disulfide bonds, Gly1), and BI–312169[24,25] (class III, 2 cysteines, 1 disulfide bond, Gly1). The backbone of the tail regions is shown in blue, the ring forming aspartate/glutamate residues are shown in red, the backbone of the remaining macrolactam rings is shown in yellow and disulfide bonds are highlighted in green.

variant[30] were shown to be heat labile, which demonstrates that thermic stability of a lasso peptide apparently depends on the size and position of the plug amino acids and the general flexibility of its C-terminal tail.

The majority of known lasso peptides possess various biological properties, ranging from antibacterial (e.g. capistruin,[12,31] microcin J25,[2,32,33] lariatin A/B[15] and astexin–1[14]) and inhibitory (e.g. propeptin[9] and siamycin I/II[7]) to receptor antagonistic activities (e.g. anantin,[1] BI–32169,[10] and RES–701/3[4,6]). As lasso peptides are from ribosomal origin and as the enzymatic maturation machinery was shown to have a rather relaxed specificity towards substitutions of most of the residues in the lasso peptide scaffold, they are also promising candidates for epitope grafting approaches. This was recently successfully demonstrated through the incorporation of the bioactive RGD epitope in the loop region of microcin J25, yielding a potent $\alpha_v\beta_3$ integrin receptor antagonist.[34]

In general, lasso peptides can be divided into three different classes, which are discerned by the number of cysteine residues and the number of corresponding disulfide bonds. The exact criteria for this classification and a representative of each class are shown in Figure 1.

Class I and class III lasso peptides contain disulfide bonds as a major or an additional stabilizing element. Such disulfide bonds are not present in class II lasso peptides, as they solely rely on the use of sterically demanding residues for lasso fold maintenance. Known representatives of class II lasso peptides feature a wide variety of plug amino acids, ranging from tyrosine (e.g. microcin J25[20–22]) to arginine (e.g. capistruin[12]) to asparagine (e.g. lariatin[15]) and greatly differ in regard to their loop size. The sterical interactions between the plug amino acid and the macrolactam ring is shown for the examples of capistruin, caulosegnin I, lariatin, and microcin J25 in Figure 2.

Even though a total of 17 different lasso peptides are known (namely 4 class I, 12 class II, and 1 class III lasso peptides), the corresponding biosynthetic gene clusters were identified in only five instances,[12–14,35,36] while two additional gene clusters found by a mass spectrometry guided approach are assumed to be linked with lasso peptide biosynthesis.[16] The five identified gene clusters are responsible for the production of astexin–1,[14] capistruin,[12] caulosegnins I-III,[13] lariatin,[36] and microcin J25.[35] An overview of the arrangements of these clusters and their producing strains is shown in Figure 3. A comparison of all gene clusters known so far can be found in Table I.

The underlying mechanism of the biosynthesis of lasso peptides was investigated in particular with the enzymes involved in microcin J25 maturation. BLAST analysis of McjC reveals homology to asparagine synthetases, which is true for all known C proteins so far. This is why they were proposed to catalyze the activation of the aspartate/glutamate side chain and to be generally involved in the macrocyclization process.[37–39] These functions were confirmed through *in vitro* studies, which not only revealed the ATP–dependency of this enzyme, but also were able to show that McjC is capable of processing a leaderless precursor peptide into mature microcin J25.[40]

BLAST analysis of McjB and other B proteins only reveals homology of their C-terminal parts to transglutaminase core domains. As these enzymes all contain the conserved catalytic triad of cysteine proteases, it was assumed that their role in lasso peptide maturation is the cleavage of the precursor

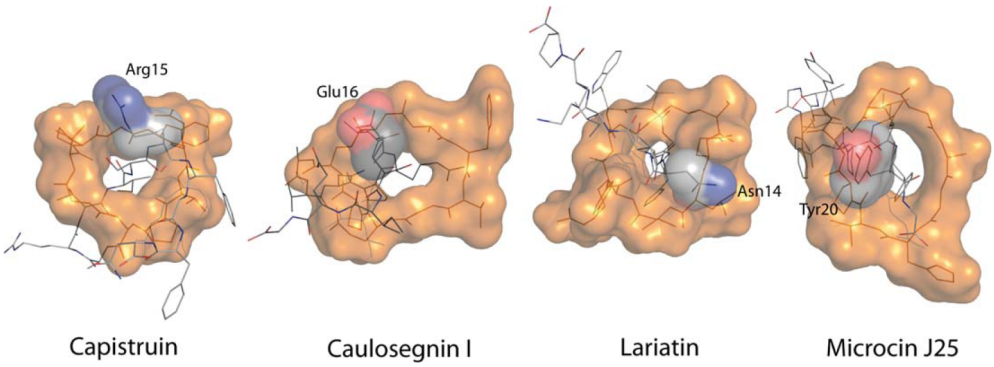


FIGURE 2 Interactions between the C-terminal tail and the macrolactam ring of capistruin,[12] caulosegnin I,[13] lariatin[15], and microcin J25.[20–22] The surface of the rings is colored in orange, the surface of the side chains of the plug amino acids positioned below the ring is colored by elements. Steric hindrance caused by the depicted side chains prevents the unthreading of the lasso peptides. All lasso peptides are shown from below the ring.

peptide.[38,39] This notion was recently confirmed by *in vitro* studies, which at the same time surprisingly revealed an ATP–dependency of McjB, suggesting a possible chaperone function of this enzyme.[40] Mutations were incorporated into *mcjB* and *mcjC* for both *in vitro*[40] and *in vivo*[41] studies to investigate residues previously suggested to be of importance for the catalytical function of these enzymes. These studies confirmed the role of C150, H182, and D194 as catalytical triad of McjB, as well as the importance of S199 and D203 for the enzymatic activity of McjC. Interestingly the *in vitro* studies also demonstrate an interdependency of McjB and McjC, shown by the lack of activity by the lone enzymes and a single catalytical activity (either macrocylization or proteolytic cleavage of the precursor) being observable when one enzyme is incubated with an inactive variant of the other enzyme.[40] A schematic representation of the proposed maturation mechanisms of lasso peptides is shown in Figure 4.

Mutational analysis of the capistruin,[30] caulosegnin,[13] and microcin J25[42,43] precursor peptides revealed the high importance of the threonine at the penultimate position of the leader peptide, which seems to be widely conserved for lasso peptide precursors.

BLAST homology searches for CapD, LarD, and McjD reveal homologies to ATP–binding cassette transporters, which are assumed to confer an immunity against the produced compounds to their host organisms.[2,12,15] This was proven in the case of McjD.[35] Interestingly, the clusters for both caulosegnin and astexin–1 production lack such a protein, while at the same time no, or only a very weak, antimicrobial activity was observed for the derived lasso peptides.[13,14]

Concerning their origin, all known lasso peptides were isolated from either actinobacteria (11 in total) or proteobacteria (6 in total). While lasso peptide production in actinobacteria seems to be most efficient in rich media,'[1,3–11,16,36] proteobacteria seem to have a preference for minimal media in regard to lasso peptide biosynthesis.[2,13]

In respect to the heterologous production in *E. coli* of lasso peptides stemming from proteobacteria, it was shown that production could be vastly increased through exchange of the intergenic region between the genes encoding the precursor peptide and the processing enzymes with an *E. coli* optimized ribosomal binding site.[14,43] Another way to improve lasso peptide yields is via control of the expression of the precursor peptide through use of a strong, inducible promoter, while the processing enzyme expression is controlled by the constitutive promoter found upstream of the *mcjBCD* gene cluster.[13,44] It was shown that clusters derived from proteobacteria also exhibit a preference for minimal media for an efficient heterologous lasso peptide biosynthesis in *E. coli.*[12–14]

In this study we apply our genome mining approach exclusively to lasso peptide biosynthetic gene clusters found in proteobacteria and combine it with heterologous production optimization approaches that were recently reported. We investigated the lasso peptide production strictly in M9 minimal media, as previous studies showed its high potency for heterologous production of proteobacterial lasso peptides.[13,14,43] Production of lasso peptides was confirmed by mass spectrometrical methods and their proposed primary structures were proven by tandem mass spectrometry. In this way we were able to identify ten novel functional lasso peptide biosynthetic gene clusters producing twelve previously unknown lasso peptides, showing the high efficiency of our approach for identification and production of such compounds from clusters found in proteobacteria.

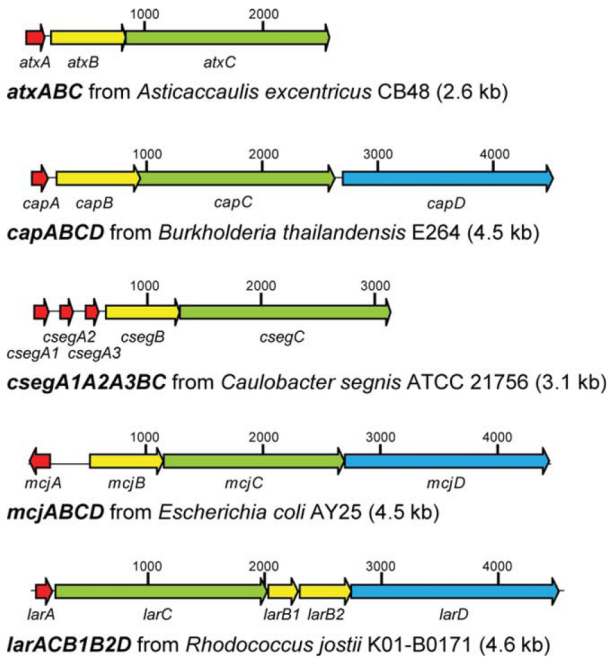


FIGURE 3 Overview of all known lasso peptide biosynthetic gene clusters and the corresponding bacterial strains. The designation of the genes from the lariatin biosynthesis cluster was changed according to the general nomenclature for lasso peptide biosynthetic gene clusters suggested by van der Donk *et al.*[50] (*larB* → *larC, larC* → *larB1, larD* → *larB2, larE* → *larD*). We propose that LarB1 and LarB2 together fulfill the same role as the larger B proteins in the other gene clusters. This notion is based on the observation that LarB2 consists mainly of a transglutaminase core domain, including the conserved motif of cysteine proteases that is also present in the C–terminal regions of the other B proteins. At the same time LarB2 lacks the N–terminal region that shares no homology to any known domains and is shortened compared to its homologs by approximately the size of LarB1 (see Table I). Taking all this into consideration with the fact that LarB1 and LarB2 are directly adjacent to each other, it is likely that they represent an example of a split B protein.

MATERIALS AND METHODS

Bacterial Strains and General Methods

Burkhholderia rhizoxinica HKI454 (DSM no. 19002), *Rhodanobacter thiooxydans* LCS2 (DSM no. 18863), *Sphingobium japonicum* UT26 (DSM no. 16413), and *Sphingopyxis alaskensis* RB2256 (DSM no. 13593) were purchased from the German Collection of Microorganisms and Cell Cultures (DSMZ). *Rubrivivax gelatinosus* IL44 (NBRC no. 100245) was purchased from the NITE Biological Resource Center (NBRC). *Caulobacter sp.* K31 was kindly provided to us by Martin Thanbichler (Max Planck Institute for Terrestrial Microbiology, Marburg), *Sphingobium yanoikuyae* XLDN2–5[45] was kindly provided to us by Ping Xu (Shanghai Jiao Tong University) and chromosomal DNA from *Phenylobacterium zucineum* HLK1[46] was kindly provided to us by Xun Hu (Zhejiang University Cancer Institute). *E. coli* TOP10 was used for cloning, *E. coli* BL21(DE3) for heterologous expression and both were purchased from Invitrogen. DNA dideoxy sequencing was done to confirm the identity of constructed plasmids and mutants thereof. Oligonucleotides and carboxypeptidase Y were purchased from Sigma Aldrich. Restriction enzymes, Phusion polymerase and T4 DNA ligase were purchased from New England Biolabs.

Genome Mining Approach

For the genome mining, the position-specific iterated basic local alignment search tool (PSI-BLAST) was applied,[47] which uses identified results above a certain score threshold for a subsequent round of BLAST analysis. For the first round McjB was used as query with the following parameters: maximum target sequences of 500, expect threshold of 10 and PSI-BLAST threshold of 0.005. BLOSUM62 algorithm was used for judgment of amino acid similarity. This first round yielded 41 putative McjB homologs above the defined threshold including many that were previously observed in other genome mining approaches.[14,38,39] The following iterations yielded 79, 112, 121 and in the fifth round 124 homologs of McjB above the defined threshold. Only 10 additional hits were found below the threshold. All hits were investigated for the presence of a McjC homolog in the vicinity of 10 kb, which revealed 32 false positive hits, where an orphan B homolog or an incomplete cluster was identified. The remaining 102 identified hits were further investigated for the presence of additional genes and small open reading frames (ORFs) that could encode possible precursor peptides, adjacent to the B and C homologs.

Cloning of the Putative Lasso Peptide Biosynthetic Gene Clusters and Incorporation of Ribosomal Binding Sites

For cloning, the identified gene clusters were amplified from respective genomic DNA samples by PCR using fitting oligonucleotide primers (Supporting Information Table S1) and Phusion polymerase. After purification of the DNA amplicons by agarose

Table I Comparison of the Proteins Encoded by all Known Lasso Peptide Biosynthetic Gene Clusters

Cluster	A protein	B protein	C protein	D protein	Produced Lasso Peptides
atxABC	51 aa	209 aa	572 aa	—	Astexin-1
capABCD	46 aa	241 aa	582 aa	606 aa	Capistruin
csegA1A2A3BC	41/37/37 aa	216 aa	617 aa	—	Caulosegnin I-III
larACB1B2D	46 aa	84 aa + 147 aa	604 aa	592 aa	Lariatin
mcjABCD	58 aa	208 aa	513 aa	580 aa	Microcin J25

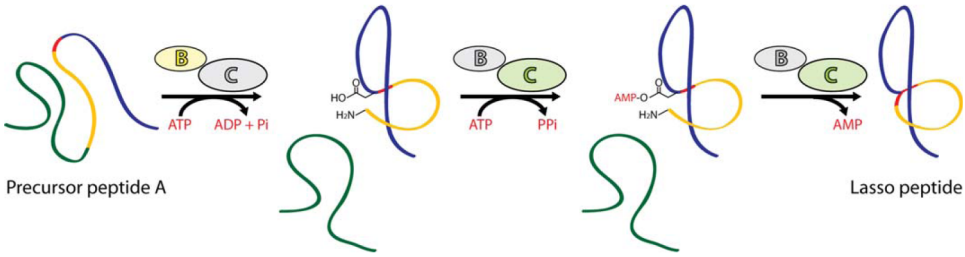


FIGURE 4 Proposed mechanism of the lasso peptide maturation.[40] In the first step the B protein cleaves off the leader sequence of the precursor peptide A and is suggested to facilitate the prefolding of the linear peptide under ATP consumption. ATP is furthermore needed for the aspartate/glutamate side chain activation catalyzed by the C protein and the subsequent macrocyclization reaction yielding mature lasso peptide. As the enzymes are interdependent, the catalyzed reactions only occur when both enzymes are present. The order of the enzymatic reactions is still unknown and thus an alternative reaction sequence is not inconceivable.

gel electrophoresis, they were digested with the stated restriction enzymes and cloned into the pET41a vector that was likewise digested. Incorporation of the ribosomal binding site was performed via PCR and use of 5'–phosphorylated oligonucleotide primers that were carrying an *E. coli* optimized ribosomal binding site sequence as overhang of the respective forward primers and were flanking the intergenic regions between precursor peptide and processing enzyme encoding genes (Supporting Information Table S2). Ligation, transformation, plasmid preparation and sequencing were done as previously described.[13] For heterologous expression, respective plasmids were retransformed into *E. coli* BL21(DE3).

Heterologous Expression

Overnight cultures were prepared in LB medium containing 50 μg/mL kanamycin at 37°C through inoculation with *E. coli* BL21(DE3) cells carrying the respective plasmids. Subsequently, 600 mL of M9 minimal medium [17.1 g/L Na_2HPO_4 2 H_2O, 3 g/L KH_2PO_4, 0.5 g/L NaCl, 1 g/L NH_4Cl, 1 mL/L $MgSO_4$ solution (2 M), 0.2 mL/L $CaCl_2$ solution (0.5 M), pH = 7.0, after autoclaving 10 mL/L glucose solution (40% w/v) and 0.2 mL/L vitamin mix (Supporting Information Table S3) were added] were inoculated to an OD_{600} of 0.01 and expression for either 1 day at 37°C or 3 days at 20°C was carried out as described in previous works.[13] Before harvesting the cells, the OD_{600} was measured and compared to the OD_{600} of all other fermentations so it could be ruled out that a possible toxicity of the produced compound hindered the cell growth and thus caused a low yield or no detectable lasso peptide production. Cells were harvested by centrifugation and extraction of the cell pellets was done with 50 mL of MeOH. Solvent was removed at 40°C and reduced pressure. Dried extracts were resuspended in a total of 800 μL of 50% MeOH, cleared by centrifugation and analyzed via LC–MS. Test fermentations for preparative HPLC purification were performed in 1 L of M9 minimal medium under the optimal conditions for the respective constructs. The cell pellet of each 1 L test fermentation was extracted with 100 mL of MeOH and the solvent was evaporated at 40°C and reduced pressure. Dried extracts were resuspended in 10 mL of 50% MeOH, cleared by centrifugation and filtration and then subjected to preparative HPLC.

Mass Spectrometric Analysis

Mass spectrometric analysis of extracts was performed with a LTQ–FT ultra instrument (Thermo Fisher Scientific) connected to a microbore 1100 HPLC system (Agilent) using 100 μL extract for each measurement. For detection the UV absorption at 215 nm was recorded. Separation was achieved using a CC 125/2 Nucleosil 300-8 C18 column (Macherey-Nagel) applying the following gradient of water/0.1% trifluoroacetic acid (solvent A) and MeCN/0.1% trifluoroacetic acid (solvent B) at a column temperature of 40°C and a flow rate of 0.2 mL/min: holding 2% B for 2 min, followed by a linear increase from 2% to X% B in 18 min, a subsequent linear increase from X% to 95% B in 15 min and holding 95% B for additional 2 min. X = 20 was chosen for zucinodin, sphingonodin I, sphingonodin II, and sphingopyxin II, while X = 30 was applied for all other compounds. Fragmentation studies and production quantification was done as previously described.[13] In cases where no UV peak was observable, quantification was done by comparing the areas of the combined extracted ion peaks of singly and doubly charged compounds from wild type to optimized constructs.

Purification

Lasso peptides were purified by preparative HPLC using a microbore 1100 HPLC system (Agilent) with a VP 250/21 Nucleodur C18 Htec 5 μm column (Macherey-Nagel). For detection, the UV absorption at 215 nm was recorded. Crude pellet extracts were subjected to the following gradient of water/0.045% formic acid (solvent A) and MeOH/0.05% formic acid (solvent B) at room temperature and a flow rate of 18 mL/min: Linear increase from 20 to 30% B in 5 min, followed by a linear increase from 30 to 95% B in 25 min and holding 95% B for additional 3 min. Observed retention times were around 19 min for caulonodins I–III, 21 min for rubrivinodin, 23 min for sphingonodin I, 20 min for sphingopyxin I, 17 min for sphingopyxin II, and 22 min for syanodin I. Lasso peptide containing fractions were pooled and solvent was evaporated at 40°C and reduced pressure.

Thermostability Assays

To assay the thermal stability of the different lasso peptides, 50 μL of the respective crude extracts were incubated for 1 h at 95°C. The samples were cooled to 4°C and subsequently analyzed with an LTQ–FT

ultra instrument (Thermo Fisher Scientific) under the conditions mentioned above. As reference, an additional 50 μL of the respective untreated extracts were measured under the same conditions. If a peak with the exact same mass as the investigated compound, but a different retention time is observed with significant intensity after the heat treatment, it most likely is due to a thermal conversion of one topology of the compound into another one. In cases of a lasso peptide this conversion would in all likelihood correspond to the unthreading of the lasso fold yielding a branched cyclic peptide.[13,30]

Carboxypeptidase Y Assays

To investigate the stability of the different lasso peptides against carboxypeptidase Y digestion, assays were performed with 1 mg/mL solutions of the purified lasso peptides and a solution of 0.1 U/mL carboxypeptidase Y, using MES buffer (50 m*M* MES, 1 m*M* $CaCl_2$, pH = 6.75) as solvent in both cases. 10 μL of the respective lasso peptide solutions were mixed with 10 μL of the carboxypeptidase Y solution and incubated for 4 h at 25°C. Subsequently, the reaction mixtures were diluted with 80 μL ddH_2O and stored at 4°C until analyzed by LC–MS. As references, 10 μL of the respective lasso peptide solutions were diluted with 90 μL ddH_2O. These experiments were repeated for the heat labile lasso peptides after incubating them for 4 h at 95 °C.

In case of the four lasso peptides that were produced in amounts not allowing their isolation, these assays were performed using their respective pellet extracts. For each measurement 50 μL of extract was dried and subsequently solved in 10 μL MES buffer. All following procedures were done as described above.

RESULTS AND DISCUSSION

B Protein-Centric Genome Mining in Proteobacteria Reveals a Manifold of Novel Putative Lasso Peptide Biosynthetic Gene Clusters

In this study, our genome mining approach was applied to the identification of novel lasso peptide biosynthetic gene clusters from proteobacteria. The reasoning behind focusing exclusively on proteobacteria was not only that *E. coli* is closely related, which was deemed to be beneficial for the heterologous expression of the found clusters, but also because recent publications demonstrated how the production of lasso peptides derived from a proteobacterial gene clusters can be optimized in *E. coli*.[13,14,43,44]

Thus, the previously described B protein-centric genome mining approach[12,13,39] was applied to all available proteobacterial genomes and a selection of putative lasso peptide biosynthetic gene clusters was cloned into the pET41a expression vector and screened for lasso peptide production. The general procedure for the selection of a cluster and the subsequent test for cluster functionality is depicted in Figure 5.

The reason for choosing the B protein-centric genome mining approach over the recently reported precursor-centric

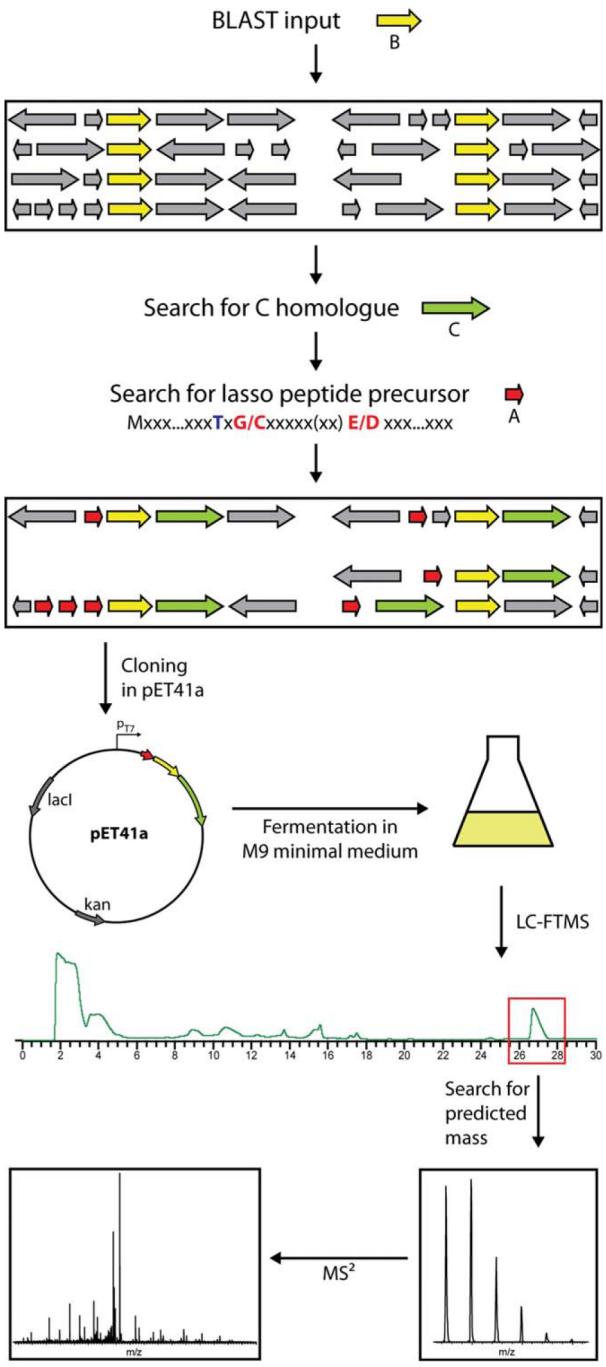


FIGURE 5 Schematic presentation of the applied genome mining approach and the subsequent confirmation of cluster functionality.

genome mining approach by Link and coworkers[14] is that B protein homologs are found almost exclusively in lasso peptide biosynthetic gene clusters and thus the resulting hits of a BLAST homology search of known B proteins would yield predominantly promising clusters, while the search for small

ORFs fitting the criteria for lasso peptide precursors would produce much wider spread results and would have a much lower hit rate. For example in the recent study by Link and coworkers (which included all available genomic data and was not solely focused on proteobacteria) 336,425 ORFs, fitting their applied precursor criteria, were initially identified, but only 98 ORFs, which were located in 79 gene clusters, remained upon closer inspection.[14] Another advantage of our approach is that clusters lacking or featuring an uncommon precursor would also be identified, while they would be overlooked by a precursor-centric approach. Of course a limitation of this method is that every cluster has to be checked manually for presence and suitability of a precursor and thus it cannot be automated as easily, making the process more time consuming and a more effective preselection necessary. This is why the initial focus on search for B homologs was deemed to be the most suitable procedure.

Applying our genome mining approach we were able to identify 102 different gene clusters from a total of 87 different proteobacterial strains. An overview over all clusters identified is shown in Figure 6, while Supporting Information Tables S4 and S5 provide further information about these clusters and all corresponding precursor peptides.

Interestingly, clusters with an A B C D arrangement are not as widespread in proteobacteria, as it was anticipated on the basis that the first two examples of proteobacterial lasso peptide biosynthetic gene clusters, which produce microcin J25 and capistruin, both feature such an organization.[12,35] In contrast, this overview suggests that the presence of an exporter, and maybe even the corresponding antimicrobial activity of the produced compound, is a rather uncommon occurrence in the ordinary kind of lasso peptide biosynthetic gene clusters found in proteobacteria and is predominantly confined to *Burkholderia* species. A much larger group of identified gene clusters consists of a simple A B C arrangement lacking an ATP–binding cassette transporter as in the clusters of *C. segnis* and *A. excentricus*.[13,14] In this regard it is an interesting observation that most of the A B C clusters are preceded by a GntR–type regulator protein and succeeded by an inversely orientated gene cluster that at least contains a TonB–dependent receptor like protein, which often is accompanied by genes encoding a transcriptional regulator, an anti-FecI sigma factor FecR homolog and a putative peptidase. This motif is not found for any other cluster organization in proteobacteria.

Furthermore, our study reveals an intriguing subgroup of putative lasso peptide biosynthetic gene clusters that all feature a split B protein similar to the one found in the lariatin biosynthesis cluster[36] and often also contain a putative transporter protein. Interestingly, most of these clusters also comprise a gene encoding a protein with homology to Hpr kinases, which are predominantly situated upstream of their respective C protein encoding gene. Upon closer inspection, some of these clusters also contain other additional genes in their arrangements, with multiple instances of clusters featuring proteins with sulfotransferase homology. At the present time there are no known functional lasso peptide biosynthetic gene clusters in which similar enzymes are found, so their function currently remains elusive. Nonetheless, it is not unreasonable to assume that their function is involved in the catalysis of novel post-translational modifications or is of a regulatory nature.

In regard to suitable lasso peptide precursors, a total of 71 out of 98 precursor peptides situated in 61 different gene clusters were identified that conform with the established criteria for lasso peptide precursors. The putative lasso peptides predicted from their amino acid sequences exclusively belong to the group of class II lasso peptides, giving rise to the notion that class I and class III lasso peptides may only be found in actinobacteria. In terms of ring size and ring forming amino acids, aspartate appears to be preferred over glutamate with a total of 53 instances where an aspartate is found at position 8 or 9, while glutamate residues at this positions were only found in 23 cases (as 5 of the predicted lasso peptides possess such residues at both position 8 or 9, the total number of possible ring forming amino acids exceeds the total number of proposed precursor peptides). From these aspartate residues, 21 would form an eight–membered and 32 would form a nine–membered macrolactam ring. This rather even distribution is not found for glutamate residues, which show a clear preference for forming an eight–membered macrolactam ring, with 18 instances of a glutamate at position 8 compared to only 5 cases where a glutamate is found at position 9.

Interestingly, there are also three examples where the only suitable amino acid for ring formation is a glutamate residue at position 7 of the lasso peptide sequence. Closer inspection reveals that a total of 7 lasso peptides possess a glutamate at this position, whereas aspartate is not found at this position in any of the precursors. Even though the presence of a seven–membered macrolactam ring is not completely inconceivable, there is currently no known example of a lasso peptide with such a small ring. A summary of how often a ring forming amino acid is found at which position of a predicted lasso peptide is shown in Table II.

In addition, a couple of clusters without a precursor fitting the general criteria (or no putative precursor at all) were identified. Enzymes of these clusters may have a different recognition motif than the enzymes of all currently known lasso peptide biosynthetic gene clusters. This could explain why their precursor peptides differ from the motif assumed to be common to all lasso peptide precursors or why no putative precursor peptide was found at all.

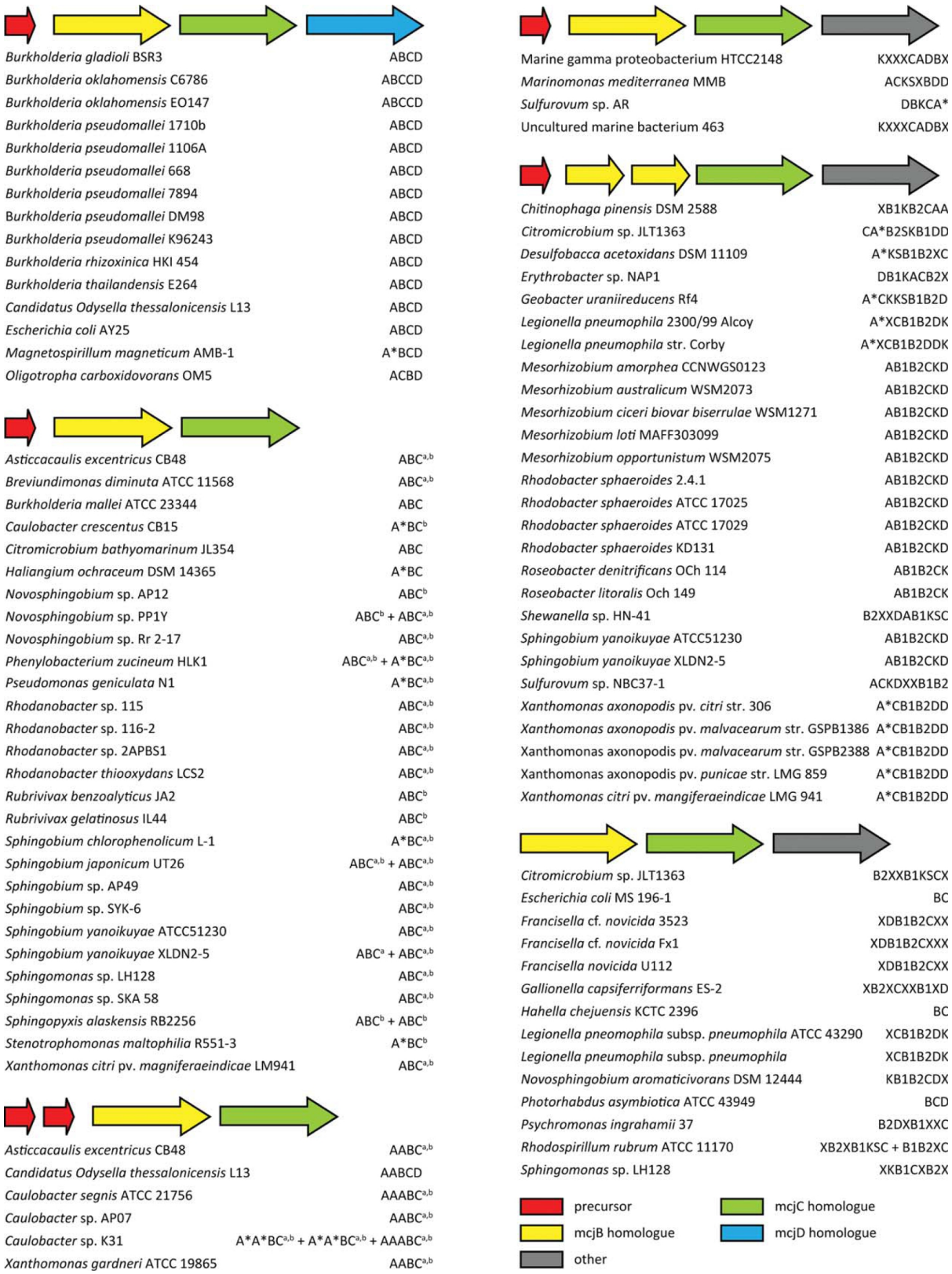


FIGURE 6 Results of the B protein-centric genome mining in proteobacteria. Clusters are ordered according to their arrangements. Precursors marked with an asterisk do not fit the general criteria for lasso peptide precursors, lacking either a threonine residue at the penultimate position of the leader peptide or a glycine/cysteine residue at the first position of the lasso peptide. Clusters marked with **a** are preceded by a gene encoding a GntR–type regulator protein, clusters marked with **b** are succeeded by an inversely orientated gene cluster containing a TonB–dependent receptor. **K** is used as an abbreviation for kinase homologs, **S** for sulfotranferase homologs and **X** is used for proteins of unknown function.

Table II Number of Instances Where Either an Aspartate or Glutamate Residue Occurs at Position 7, 8, or 9 of a Predicted Lasso Peptide Derived From the Clusters With a Suitable Precursor Peptide

Amino Acid	Position 7	Position 8	Position 9	Total No. of Occurrences
Aspartate	0	21	32	53
Glutamate	7	18	5	30

Cloning of Putative Lasso Peptide Biosynthetic Gene Clusters for Heterologous Expression in *E. coli* BL21(DE3)

From all identified putative lasso peptide biosynthetic gene clusters a total of 10 clusters from eight different organisms were chosen for further investigation. These clusters were cloned into the pET41a expression vector and the intergenic region between their precursor gene(s) and the genes encoding their processing enzymes was replaced with a strong *E. coli* ribosomal binding site (RBS). In case of *C.* sp. K31, where more than one precursor was present, several RBS containing constructs were created, each containing only one of its respective precursors, as in case of the caulosegnin multi-precursor system this modification was shown to be beneficial for the overall yield.[13] A comparison of all the selected clusters can be found in Table III.

As the incorporation of an *E. coli* optimized RBS was proven to be an efficient way to significantly increase the lasso peptide yields in the astexin–1,[14] capistruin,[43] and caulosegnin[13] systems, clusters for further investigation were chosen based on their similarity to these systems. Additionally, the cluster from *B. rhizoxinica* HKI454 was cloned without the gene encoding its putative exporter protein. This was done to allow for the isolation and identification of the produced compound from pellet extracts like for the other systems.

From the chosen organisms two contain two different lasso peptide biosynthetic gene clusters, namely *S. japonicum* UT26 and *S. alaskensis* RB2256. Furthermore, *S. yanoikuyae* XLDN2–5 contains three clusters, from which two are positioned at the end of a sequencing contig, which is why only parts of their genes could be identified and thus they were not further examined. Interestingly, one of the partially sequenced clusters is highly similar to the one investigated and thus most likely shares its A B C organization, while the third one shows vast differences. Very recently, another strain of this subspecies (*S. yanoiukyae* ATCC51230) was sequenced, which features almost identical copies of the found genes from the third *S. yanoikuyae* XLDN2–5 cluster and thus the complete arrangement of this cluster can be deduced. Interestingly, it features a kinase homolog in between its C and D protein as well as a split B protein, similar to what has been shown in the lariatin system.[36] This organization resembles clusters found in *Mesorhizobia* and *Rhodobacter* species and BLAST analysis of the respective genes from *S. yanoikuyae* XLDN2–5 and *S. yanoikuyae* ATCC51230 reveals a high similarity between these kind of clusters from all three species, suggesting that they could stem from the same origin.

It is also worth mentioning that the investigated cluster of *C.* sp. K31 features three precursor peptides, similar to the cluster from *C. segnis*. While the first two precursor peptides of the *C.* sp. K31 cluster are highly similar, the third precursor features not only a more diverse primary structure, but also contains a lone cysteine residue. This appears to be rather rare in lasso peptides and is only found in three other precursor peptides. As such the lasso peptide derived from *C. sp.* K31 A3 would be the first example of a lasso peptide with an uneven number of cysteines.

Even though three different putative lasso peptide biosynthetic gene clusters, encoding a total of seven precursor peptides, were identified in the genome of *C.* sp. K31, only one cluster was further examined, because the others contain precursor peptides not in accordance with the general criteria for lasso peptide precursors. The same reasoning applies to the second cluster found in the genome of *P. zucineum* HLK1 and explains why it was also not considered for further investigation.

A comparison of all apposite precursor peptides which stem from one single organism is depicted in Supporting Information Figure S1, while Supporting Information Table S6 shows the similarities between different processing enzymes in clusters with A B C arrangements present in the same host organism.

Heterologous Expression and Analysis of the Produced Compounds by Mass Spectrometry

E. coli BL21(DE3) cells containing the plasmids featuring both the wild type and the modified gene clusters were subjected to fermentation in M9 minimal media under different growth conditions. Cell pellet extracts were analyzed for lasso peptide presence by mass spectrometry. In all cases a compound with the predicted mass or truncated variants thereof could be detected. An overview of all precursor peptides and the corresponding lasso peptides found as main products of their heterologous expression in *E. coli* is shown in Figure 7. The predicted primary structures and the size of the macrolactam rings were confirmed by tandem mass spectrometry (see Supporting Information Figures S2–13). For all fermentations the observed OD_{600} values for control and recombinant strains were in the same range, suggesting that none of the produced lasso peptides had a toxic effect on *E. coli* BL21(DE3). Best

Table III Comparison of the Investigated Gene Clusters

Strain	Cluster Organization	Size of Protein A/GenBank Accession No./Gene Designation	Size of Protein B/GenBank Accession No./Gene Designation	Size of Protein C/GenBank Accession No./Gene Designation	Size of Protein D/GenBank Accession No./Gene Designation
Burkholderia rhizoxinica HKI454	A B C D	50 aa YP_004028910.1 *burhA*	276 aa YP_004028969.1 *burhB*	580 aa YP_004028968.1 *burhC*	599 aa YP_004028967.1 *burhD*
Caulobacter sp. K31	A A A B C	44aa, 44 aa, 44 aa YP_001676628.1 YP_001676627.1 – *ck31_A1, ck31_A2, ck31_A3*	219 aa YP_001676626.1 *ck31_B*	579 aa YP_001676625.1 *ck31_C*	—
Phenylobacterium zucineum HLK1	A B C	39 aa – *pzucA*	212 aa YP_002131237.1 *pzucB*	571 aa YP_002131236.1 *pzucC*	—
Rhodanobacter thiooxydans LCS2	A B C	54 aa – *rhotA*	222 aa ZP_10205116.1 *rhotB*	584 aa ZP_10205117.1 *rhotC*	—
Rubrivivax gelatinosus IL44	A B C	43aa YP_005435543.1 *rugeA*	241 aa YP_005435542.1 *rugeB*	617 aa YP_005435541.1 *rugeC*	—
Sphingobium japonicum UT26	A B C	39 aa – *sjap1_A*	218 aa YP_003546448.1 *sjap1_B*	583 aa YP_003546449.1 *sjap1_C*	—
Sphingobium japonicum UT26	A B C	46 aa YP_003547071.1 *sjap2_A*	217 aa YP_003547070.1 *sjap2_B*	581 aa YP_003547069.1 *sjap2_C*	—
Sphingobium yanoikuyae XLDN2 5	A B C	44 aa – *syan1_A*	220 aa ZP_09906533.1 *syanB*	581 aa ZP_09906534.1 *syanC*	—
Sphingopyxis alaskensis RB2256	A B C	44 aa – *sala1_A*	213 aa YP_617574.1 *sala1_B*	570 aa YP_617573.1 *sala1_C*	—
Sphingopyxis alaskensis RB2256	A B C	44 aa – *sala2_A*	219 aa YP_617642.1 *sala2_B*	578 aa YP_617641.1 *sala2_C*	—

production was predominantly observed, with the clusters from *B. rhizoxinica* HKI454 and *R. gelatinosus* IL44 being the only exceptions, when the cells were incubated at 20°C over the course of 3 days. Comparison of the lasso peptide production of the unmodified gene clusters to those of modified plasmids revealed that the incorporation of the RBS greatly increased the production in almost all cases. The only exception for this is the cluster from *R. thiooxydans* LCS2, for which the produced amounts of lasso peptide of both the wild type and the modified construct were practically identical. Furthermore, the lasso peptide derived from the cluster of *P. zucineum* HLK1 and two lasso peptides derived from the cluster of *C.* sp. K31 were only detectable when using the modified constructs containing the *E. coli* optimized RBS.

The remarkable 84.5–fold increase of production for the first lasso peptide from the cluster of *C. sp.* K31 is most likely

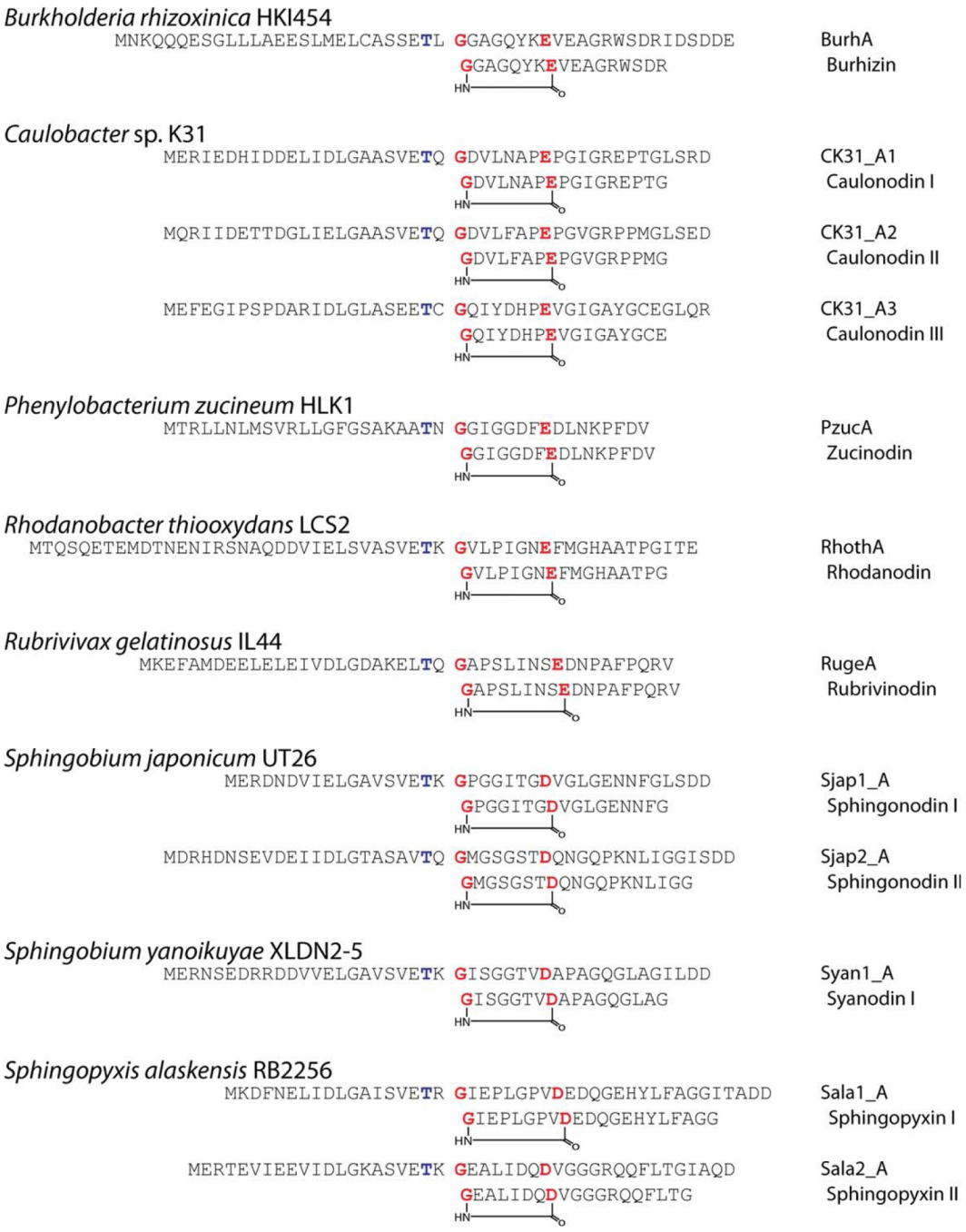


FIGURE 7 Overview of all 12 investigated lasso peptide precursor peptides and the strains they were found in. The threonine at the penultimate position of the leader peptide is marked in blue, while both amino acids involved in macrolactam ring formation are highlighted in red. The lasso peptides that were detected as main products of the heterologous expression in *E. coli* are shown below their corresponding precursor peptides. On the right the names of the precursor peptides and the names of the corresponding lasso peptides are shown, respectively.

Table IV Results of the Heterologous Expression of the Investigated Lasso Peptide Biosynthetic Gene Clusters

Lasso Peptide[a]	Precursor	Full Length	Length of the Main Product	Ring Forming aa	Ringsize	Optimal Production at	Production Increase Through RBS[b]	Heat Stable?	Yield
Burhizin	BurhA	23 aa	17 aa	Glu8	8 aa / 26 atoms	1 d, 37°C	12.4-fold	yes	nd
Caulonodin I	CK31_A1	21 aa	17 aa	Glu8	8 aa / 26 atoms	3 d, 20°C	84.5-fold	yes	3.4 mg/L
Caulonodin II	CK31_A2	21 aa	17 aa	Glu8	8 aa / 26 atoms	3 d, 20°C	nd	yes	2.2 mg/L
Caulonodin III	CK31_A3	21 aa	17 aa	Glu8	8 aa / 26 atoms	3 d, 20°C	nd	yes	0.5 mg/L
Zucinodin	PzucA	16 aa	16 aa	Glu8	8 aa / 26 atoms	3 d, 20°C	nd	yes	nd
Rhodanodin	RhotA	20 aa	17 aa	Glu8	8 aa / 26 atoms	3 d, 20°C	no increase	no	nd
Rubrivinodin	RugeA	18 aa	18 aa	Glu9	9 aa / 29 atoms	1 d, 37°C	25.0-fold	no	0.5 mg/L
Sphingonodin I	Sjap1_A	21 aa	17 aa	Asp8	8 aa / 25 atoms	3 d, 20°C	1.8-fold	yes	0.9 mg/L
Sphingonodin II	Sjap2_A	23 aa	19 aa	Asp8	8 aa / 25 atoms	3 d, 20°C	2.5-fold	no	nd
Syanodin I	Syan1_A	21 aa	17 aa / 16 aa	Asp8	8 aa / 25 atoms	3 d, 20°C	8.9-fold	no	5.2 mg/L
Sphingopyxin I	Sala1_A	26 aa	21 aa	Asp9	9 aa / 28 atoms	3 d, 20°C	7.6-fold	no	3.4 mg/L
Sphingopyxin II	Sala2_A	23 aa	19 aa	Asp8	8 aa / 25 atoms	3 d, 20°C	14.4-fold	yes	0.4 mg/L

nd = not determined.

[a] Nodin was derived from *nodus*, the latin word for knot.

[b] Production under the best condition was compared between wild type and RBS construct (see Supporting Information Table S7). This was done through comparison of the peak integrals of either the UV peaks (if possible) or extracted ion peaks of the mass signals (when overall production was too low to observe an UV signal).

not solely due to the incorporation of the RBS sequence, but also may be due to the deletion of the other two precursor encoding genes inherent to this cluster. Such a separation of genes encoding precursor peptides was previously shown to increase the overall yields in the caulosegnin system.[13] A summary of the produced compounds, the best conditions for their production and their features can be found in Table IV, alongside names that were assigned to these novel lasso peptides to simplify further discussion.

Based on UV peak intensity a moderate to good production was observed for all lasso peptides except burhizin, rhodanodin, sphingonodin II, and zucinodin, from which zucinodin was only detectable by MS. To quantify the production of the other lasso peptides, 1 L fermentations were conducted and extracts were applied to a preparative HPLC. The observed yields are listed in Table IV, ranging from 0.4 mg/L for sphingopyxin II to 5.2 mg/L for syanodin I. The high yields for caulonodin I, sphingopyxin I, and syanodin I are especially remarkable, as previous reports of optimized heterologous production of lasso peptides in *E. coli* led to amounts below 2 mg/L. The only exception for this is the system producing microcin J25, which originated from *E. coli* AY25.

To test the stability of all reported lasso peptides against thermal degradation, the extracts containing the major amounts of each lasso peptide were incubated for 1 h at 95°C and compared with untreated samples. After LC-MS analysis, the occurrence of a new peak with the exact same mass but a different retention time refers in all likelihood to the generation of a topologically different form of the examined compound. This new peak is most likely caused by thermal unthreading of the lasso fold yielding a branched cyclic peptide.[13,30] The formation of such a peak with significant intensity was observed for five of the twelve tested lasso peptides, namely rhodanodin, rubrivinodin, sphingonodin II, sphingopyxin I, and syanodin I (Supporting Information Figure S14).

These newly formed derivatives were also analyzed by tandem mass spectrometry to confirm their primary structures and thus their identities. A comparison of the fragmentation spectra (Supporting Information Figure S15) reveals that the fragmentation behavior of rhodanodin, sphingonodin II, and syanodin I undergoes drastic changes with regard to the relative signal intensities of their different fragments when they are unthreaded. In contrast, rubrivinodin and sphingopyxin I show an almost identical fragmentation pattern in lasso and in branched cyclic topology. The causes for these different behaviors remain unclear. On the one hand, it could be reasoned that, compared to a branched cyclic topology, a very rigid lasso fold could alter the locations of preferred bond breakages during fragmentation due to the induced conformational stress. On the other hand, this would also imply that the fragmentation pattern would align more and more to the one of a branched cyclic peptide the more relaxed the lasso fold gets. However, a final conclusion in this regard will await the structure elucidation of these compounds.

To further prove that the presented twelve compounds are true lasso peptides, carboxypeptidase Y assays were performed. In case of a lasso peptide, the degradation of the C-terminal

peptide chain by carboxypeptidase Y would stop as soon as the protease reaches close proximity to the part of the tail that is shielded by the surrounding macrolactam ring. Alteration of the topology to a branched cyclic peptide would therefore cancel this effect, allowing the carboxypeptidase Y to further degrade the peptide's tail. A direct comparison of a carboxypeptidase Y digestion of a lasso peptide with the digestion of its branched cyclic analog would thus result in a shorter main product of the latter.

In this way, the eight isolated lasso peptides were assessed through treatment with carboxypeptidase Y. No truncation was observed for caulonodin I, caulonodin III and rubrivinodin, while all other lasso peptides lost no more than one (sphingonodin I) or two (caulonodin II, sphingopyxin I, sphingopyxin II and syanodin I) amino acids. For comparison, the heat labile lasso peptides rubrivinodin, sphingopyxin I and syanodin I were unthreaded by prolonged exposure to 95°C and subsequently treated with carboxypeptidase Y. As expected, the found identified products in all three cases were shorter than in assays without prior heat treatment, with rubrivinodin and sphingopyxin I now lacking five, while syanodin I was lacking three amino acids. A graphical summary of these assays can be found in the Supporting Information (Supporting Information Figure S16).

For investigation of the low yield lasso peptides, the carboxypeptidase Y assays were performed using the respective pellet extracts. This revealed that the heat stable lasso peptides burhizin and zucinodin were almost completely resistant against carboxypeptidase Y degradation, which underscores their proposed topology (Supporting Information Figure S17). Investigation of the susceptibility of heat labile lasso peptides to carboxypeptidase Y digestion before and after heat treatment becomes much more complex when not using pure compounds. The reason for this is that the degradation products do not appear as single, distinct peaks in the UV spectrum of the chromatogram. Instead, their detection demands a thorough search for every putative truncation variant up to the macrolactam ring. Still, both assays for rhodanodin and sphingonodin II provide satisfying evidence for their proposed topologies before and after incubation at 95°C. In the case of sphingonodin II, the truncation of the C-terminal tail can proceed up to two amino acids farther when unthreaded (Supporting Information Figure S18). Unfortunately, we were unable to detect the final degradation product of the carboxypeptidase Y digestion of branched cyclic rhodanodin under the experimental conditions. However, we were able to show that while residual lasso peptide, which remained after the heat treatment, behaves identical to the unheated sample, the signals corresponding to the branched cyclic peptide completely vanish after incubation with carboxypeptidase Y (Supporting Information Figure S19), proving its high susceptibility for proteolytic degradation.

In summary, we were able to produce all twelve predicted lasso peptides through incorporation of an *E. coli* optimized RBS between the precursor peptide and the processing enzyme encoding genes, eight of which are produced in amounts sufficient for isolation. From these twelve lasso peptides seven show high heat stability, while the other five are easily unthreaded upon exposure to temperatures of 95°C. In all cases, carboxypeptidase Y treatment provided further evidence for their proposed lasso topology.

Comparison of All Functional Processing Enzymes

The expansion of known functional lasso peptide biosynthetic gene clusters from five to fifteen in this study allowed us to explore some general characteristic features and conserved motifs of the associated processing enzymes. For this, an alignment of all processing enzymes was performed using the MEME algorithm and is shown in a graphical representation in Figure 8.

The identified motifs are in general accordance with those reported by Link and coworkers,[14] which were discovered by aligning the processing enzymes from the clusters producing microcin J25 and capistruin with nine additional putative lasso peptide biosynthetic gene clusters (including three clusters shown to be functional in this work). Nonetheless, the fourth motif for protein C was not reported before. As expected, the second motif of the B protein includes the catalytic triad of a cysteine protease, while the first motif of the C protein contains the Ser and Asp residues, which were recently shown to be essential for catalytic activity of McjC.[41,42]

As suspected LarB1 contains only the first motif of the B proteins, while LarB2 features only the second one. This further supports our hypothesis that the lariatin biosynthetic gene cluster has a split B protein, in which enzymatic functionality of the N-terminal region found in other B proteins is present in LarB1, while LarB2 retains the protease functionality as it contains the catalytic triad characteristic for cysteine proteases.

CONCLUSION

In this work, we demonstrate the high efficiency and broad applicability of combining the B protein–centric genome mining approach for identifying functional lasso peptide biosynthetic gene clusters in proteobacteria with an optimized heterologous expression strategy in *E. coli*. Through this, we were able to identify twelve novel lasso peptides, which were detected via high–resolution Fourier transform mass

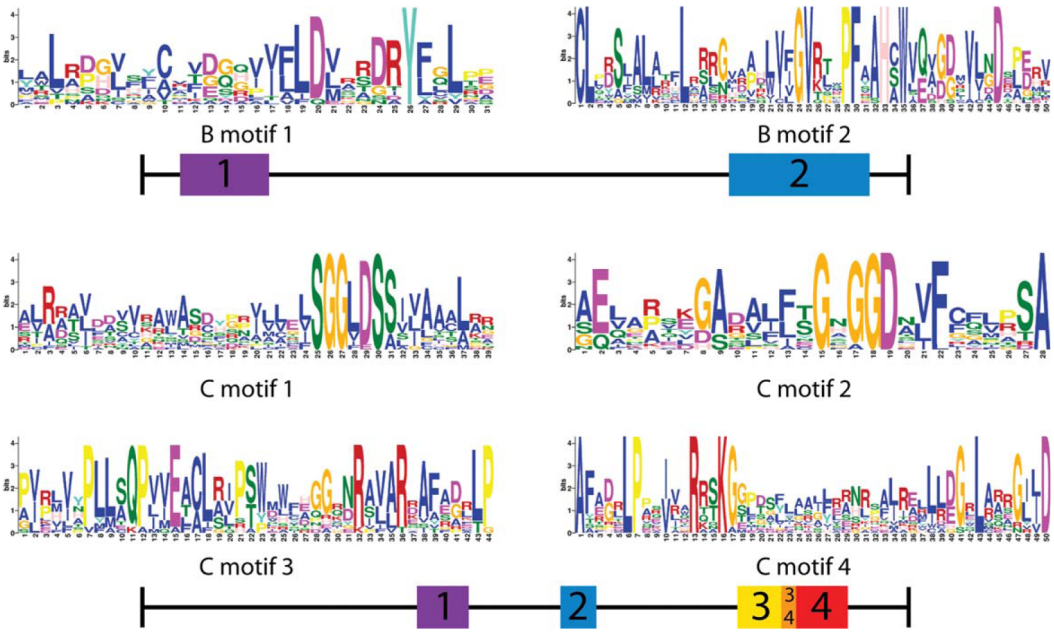


FIGURE 8 Common motifs in the lasso peptide processing enzymes B and C identified through comparison of all 15 known functional lasso peptide biosynthesis clusters using the MEME algorithm (see Supporting Information Tables S8–S13). Location of different motifs in each enzyme is shown approximately in a scale representation. The slight overlap of motifs 3 and 4 of the C protein is depicted schematically through mixing of the motifs color codes.

spectrometry and whose predicted primary structures were confirmed by tandem mass spectrometry.

The results of our genome mining study in proteobacteria revealed that in this phylum the presence of genes encoding an ATP–binding cassette transporter is rather uncommon in ordinary lasso peptide biosynthetic gene clusters and that the predominant cluster arrangement appears to be A B C, preceded by a GntR–type regulator protein and succeeded by an inversely orientated gene cluster containing numerous conserved genes, including a TonB–dependent receptor protein. The fact that in all available proteobacterial genomic data not a single instance of a class I or III lasso peptide was found, as well as the observation that only 4 out of 98 predicted lasso peptides contained a cysteine residue at all, suggests that these classes of lasso peptides could be exclusive to actinobacteria. Furthermore, our study shows that there is another interesting cluster arrangement present in proteobacteria, namely an A B1 B2 C K D organization (where K stands for a kinase–type protein) that in some cases is even further extended by some other proteins, e.g. sulfotransferase homologs. Based on this, future studies should focus on the investigation of non–canonical lasso peptide biosynthetic gene clusters and on the development of an efficient approach for genome mining and heterologous expression for other bacteria. In this regard, actinobacteria seem like an especially promising target, as 11 from the 17 previously known lasso peptides were found in organisms of this phylum and numerous actinobacteria carry novel putative lasso peptide biosynthetic gene clusters as shown by a recently published genome mining study done by Link and coworkers.[14]

Furthermore, we have shown the effectiveness of employing mass spectrometry for initial detection of novel lasso peptides. Of course, mass spectrometrical analysis alone cannot prove the three dimensional lasso structure of the presented compounds, but the combination of confirming their predicted primary sequences and their proposed ring sizes by mass spectrometry encourages further investigations towards their structural elucidation. The newly discovered rhodanodin, rubrivinodin, sphingonodin II, sphingopyxin I, and syanodin I show that some of these compounds can adopt another topology and further highlight that not all lasso peptides have intrinsic stability against thermal denaturation. Combining the data obtained by mass spectrometry with the results of the carboxypeptidase Y assays and the heterologous origin of all shown compounds, we have provided a solid body of evidence of these peptides being true lasso peptides. Nevertheless, the only indisputable proof of their topology is the elucidation of their structures by NMR spectroscopy or X–ray diffraction.

Finally, we note that the lasso peptides described here are complementary to the various head-to-tail cyclic peptides described elsewhere in this special issue of Biopolymers Peptide Science dedicated to circular proteins. In particular, the lasso peptides achieve the high stability properties reported for cyclotides[48,49] using an alternative topological strategy.

The authors thank Prof. Martin Thanbichler from the Max Planck Institute for Terrestrial Microbiology in Marburg for supplying us with *Caulobacter* sp. K31, Prof. Ping Xu from the Shanghai Jiao Tong University for providing *Sphingobium yanoikuyae* XLDN2–5, and Prof. Xun Hu from the Zhejiang University Cancer Institute for providing us with a chromosomal DNA sample of *Phenylobacterium zucineum* HLK1. Furthermore, they like to thank Marco Hornung, Dr. Ivana Djurdjevic, and Prof. Wolfgang Buckel for helping us with the anearobic fermentation of *Rubrivivax gelatinosus* IL44 as well as Elena Schütz, Lara Golde, and Dr. Thomas Knappe for initial help with these studies.

REFERENCES

1. Weber, W.; Fischli, W.; Hochuli, E.; Kupfer, E.; Weibel, E. K. J Antibiot (Tokyo) 1991, 44, 164–171.
2. Salomon, R. A.; Farias, R. N. J Bacteriol 1992, 174, 7428–7435.
3. Helynck, G.; Dubertret, C.; Mayaux, J. F.; Leboul, J. J Antibiot (Tokyo) 1993, 46, 1756–1757.
4. Morishita, Y.; Chiba, S.; Tsukuda, E.; Tanaka, T.; Ogawa, T.; Yamasaki, M.; Yoshida, M.; Kawamoto, I.; Matsuda, Y. J Antibiot (Tokyo) 1994, 47, 269–275.
5. Potterat, O.; Stephan, H.; Metzger, J. W.; Gnau, V.; Zähner, H.; Jung, G. Liebigs Ann Chem 1994, 7, 741–743.
6. Ogawa, T.; Ochiai, K.; Tanaka, T.; Tsukuda, E.; Chiba, S.; Yano, K.; Yamasaki, M.; Yoshida, M.; Matsuda, Y. J Antibiot (Tokyo) 1995, 48, 1213–1220.
7. Tsunakawa, M.; Hu, S. L.; Hoshino, Y.; Detlefson, D. J.; Hill, S. E.; Furumai, T.; White, R. J.; Nishio, M.; Kawano, K.; Yamamoto, S.; Fukagawa, Y; Oki, T. J Antibiot (Tokyo) 1995, 48, 433–434.
8. Yano, K.; Toki, S.; Nakanishi, S.; Ochiai, K.; Ando, K.; Yoshida, M.; Matsuda, Y.; Yamasaki, M. Bioorg Med Chem 1996, 4, 115–120.
9. Kimura, K.; Kanou, F.; Takahashi, H.; Esumi, Y.; Uramoto, M.; Yoshihama, M. J Antibiot (Tokyo) 1997, 50, 373–378.
10. Potterat, O.; Wagner, K.; Gemmecker, G.; Mack, J.; Puder, C.; Vettermann, R.; Streicher, R. J Nat Prod 2004, 67, 1528–1531.
11. Kimura, K.; Yamazaki, M.; Sasaki, N.; Yamashita, T.; Negishi, S.; Nakamura, T.; Koshino, H. J Antibiot (Tokyo) 2007, 60, 519–523.
12. Knappe, T. A.; Linne, U.; Zirah, S.; Rebuffat, S.; Xie, X.; Marahiel, M. A. J Am Chem Soc 2008, 130, 11446–11454.
13. Hegemann, J. D.; Zimmermann, M.; Xie, X.; Marahiel, M. A. J Am Chem Soc 2013, 135, 210–222.
14. Maksimov, M. O.; Pelczer, I.; Link, A. J. Proc Natl Acad Sci USA 2012, 109, 15223–15228.
15. Iwatsuki, M.; Tomoda, H.; Uchida, R.; Gouda, H.; Hirono, S.; Omura, S. J Am Chem Soc 2006, 128, 7486–7491.
16. Kersten, R. D.; Yang, Y. L.; Xu, Y.; Cimermancic, P.; Nam, S. J.; Fenical, W.; Fischbach, M. A.; Moore, B. S.; Dorrestein, P. C. Nat Chem Biol 2011, 7, 794–802.
17. Frechet, D.; Guitton, J. D.; Herman, F.; Faucher, D.; Helynck, G.; Monegier du Sorbier, B.; Ridoux, J. P.; James-Surcouf, E.; Vuilhorgne, M. Biochemistry 1994, 33, 42–50.
18. Constantine, K. L.; Friedrichs, M. S.; Detlefsen, D.; Nishio, M.; Tsunakawa, M.; Furumai, T.; Ohkuma, H.; Oki, T.; Hill, S.; Bruccoleri, R. E.; Lin, P.-F.; Mueller, L. J Biomol NMR 1995, 5, 271–286.
19. Katahira, R.; Shibata, K.; Yamasaki, M.; Matsuda, Y.; Yoshida, M. Bioorg Med Chem 1995, 3, 1273–1280.
20. Bayro, M. J.; Mukhopadhyay, J.; Swapna, G. V.; Huang, J. Y.; Ma, L. C.; Sineva, E.; Dawson, P. E.; Montelione, G. T.; Ebright, R. H. J Am Chem Soc 2003, 125, 12382–12383.
21. Rosengren, K. J.; Clark, R. J.; Daly, N. L.; Goransson, U.; Jones, A.; Craik, D. J. J Am Chem Soc 2003, 125, 12464–12474.
22. Wilson, K. A.; Kalkum, M.; Ottesen, J.; Yuzenkova, J.; Chait, B. T.; Landick, R.; Muir, T.; Severinov, K.; Darst, S. A. J Am Chem Soc 2003, 125, 12475–12483.
23. Rosengren, K. J.; Blond, A.; Afonso, C.; Tabet, J. C.; Rebuffat, S.; Craik, D. J. Biochemistry 2004, 43, 4696–4702.
24. Knappe, T. A.; Linne, U.; Xie, X.; Marahiel, M. A. FEBS Lett 2010, 584, 785–789.
25. Nar, H.; Schmid, A.; Puder, C.; Potterat, O. ChemMedChem 2010, 5, 1689–1692.
26. Xie, X.; Marahiel, M. A. Chembiochem 2012, 13, 621–625.
27. Wyss, D. F.; Lahm, H. W.; Manneberg, M.; Labhardt, A. M. J Antibiot (Tokyo) 1991, 44, 172–180.
28. Yamasaki, M.; Yano, K.; Yoshida, M.; Matsuda, Y.; Yamaguchi, K. J Antibiot (Tokyo) 1994, 47, 276–280.
29. Blond, A.; Cheminant, M.; Destoumieux-Garzón, D.; Ségalas-Milazzo, I.; Peduzzi, J.; Goulard, C.; Rebuffat, S. Eur J Biochem 2002, 269, 6212–6222.
30. Knappe, T. A.; Linne, U.; Robbel, L.; Marahiel, M. A. Chem Biol 2009, 16, 1290–1298.
31. Kuznedelov, K.; Semenova, E.; Knappe, T. A.; Mukhamedyarov, D.; Srivastava, A.; Chatterjee, S.; Ebright, R. H.; Marahiel, M. A.; Severinov, K. J Mol Biol 2011, 412, 842–848.
32. Yuzenkova, J.; Delgado, M.; Nechaev, S.; Savalia, D.; Epshtein, V.; Artsimovitch, I.; Mooney, R. A.; Landick, R.; Farias, R. N.; Salomon, R.; Severinov, K. J Biol Chem 2002, 277, 50867–50875.
33. Adelman, K.; Yuzenkova, J.; La Porta, A.; Zenkin, N.; Lee, J.; Lis, J. T.; Borukhov, S.; Wang, M. D.; Severinov, K. Mol Cell 2004, 14, 753–762.
34. Knappe, T. A.; Manzenrieder, F.; Mas-Moruno, C.; Linne, U.; Sasse, F.; Kessler, H.; Xie, X.; Marahiel, M. A. Angew Chem Int Ed Engl 2011, 50, 8714–8717.
35. Solbiati, J. O.; Ciaccio, M.; Farias, R. N.; Gonzalez-Pastor, J. E.; Moreno, F.; Salomon, R. A. J Bacteriol 1999, 181, 2659–2662.
36. Inokoshi, J.; Matsuhama, M.; Miyake, M.; Ikeda, H.; Tomoda, H. Appl Microbiol Biotechnol 2012, 95, 451–460.
37. Clarke, D. J.; Campopiano, D. J. Org Biomol Chem 2007, 5, 2564–2566.
38. Duquesne, S.; Destoumieux-Garzon, D.; Zirah, S.; Goulard, C.; Peduzzi, J.; Rebuffat, S. Chem Biol 2007, 14, 793–803.

39. Severinov, K.; Semenova, E.; Kazakov, A.; Kazakov, T.; Gelfand, M. S. Mol Microbiol 2007, 65, 1380–1394.
40. Yan, K. P.; Li, Y.; Zirah, S.; Goulard, C.; Knappe, T. A.; Marahiel, M. A.; Rebuffat, S. Chembiochem 2012, 13, 1046–1052.
41. Pan, S. J.; Rajniak, J.; Cheung, W. L.; Link, A. J. Chembiochem 2012, 13, 367–370.
42. Cheung, W. L.; Pan, S. J.; Link, A. J. J Am Chem Soc 2010, 132, 2514–2515.
43. Pan, S. J.; Rajniak, J.; Maksimov, M. O.; Link, A. J. Chem Commun (Camb) 2012, 48, 1880–1882.
44. Pan, S. J.; Cheung, W. L.; Link, A. J. Protein Expr Purif 2010, 71, 200–206.
45. Gai, Z.; Wang, X.; Tang, H.; Tai, C.; Tao, F.; Wu, G.; Xu, P. J Bacteriol 2011, 193, 6404–6405.
46. Luo, Y.; Xu, X.; Ding, Z.; Liu, Z.; Zhang, B.; Yan, Z.; Sun, J.; Hu, S.; Hu, X. BMC Genomics 2008, 9, 386.
47. Altschul, S. F.; Madden, T. L.; Schaffer, A. A.; Zhang, J.; Zhang, Z.; Miller, W.; Lipman, D. J Nucleic Acids Res 1997, 25, 3389–3402.
48. Rosengren, K. J.; Daly, N. L.; Harvey, P. J.; Craik, D. J. Biopolymers (Pept Sci) 2013, 100, 453–460.
49. Poth, A. G.; Chan, L. Y.; Craik, D. J. Biopolymers (Pept Sci) 2013, 100, 480–491.
50. Arnison, P. G.; Bibb, M. J.; Bierbaum, G.; Bowers, A. A.; Bugni, T. S.; Bulaj, G.; Camarero, J. A.; Campopiano, D. J.; Challis, G. L.; Clardy, J.; Cotter, P. D.; Craik, D. J.; Dawson, M.; Dittmann, E.; Donadio, S.; Dorrestein, P. C.; Entian, K. D.; Fischbach, M. A.; Garavelli, J. S.; Goransson, U.; Gruber, C. W.; Haft, D. H.; Hemscheidt, T. K.; Hertweck, C.; Hill, C.; Horswill, A. R.; Jaspars, M.; Kelly, W. L.; Klinman, J. P.; Kuipers, O. P.; Link, A. J.; Liu, W.; Marahiel, M. A.; Mitchell, D. A.; Moll, G. N.; Moore, B. S.; Muller, R.; Nair, S. K.; Nes, I. F.; Norris, G. E.; Olivera, B. M.; Onaka, H.; Patchett, M. L.; Piel, J.; Reaney, M. J.; Rebuffat, S.; Ross, R. P.; Sahl, H. G.; Schmidt, E. W.; Selsted, M. E.; Severinov, K.; Shen, B.; Sivonen, K.; Smith, L.; Stein, T.; Sussmuth, R. D.; Tagg, J. R.; Tang, G. L.; Truman, A. W.; Vederas, J. C.; Walsh, C. T.; Walton, J. D.; Wenzel, S. C.; Willey, J. M.; van der Donk, W. A. Nat Prod Rep 2012, 30, 108–160.

3.4 Xanthomonins I-III: A New Class of Lasso Peptides With a Seven-Residue Macrolactam Ring

Julian D. Hegemann, Marcel Zimmermann, Shaozhou Zhu, Holger Steuber, Klaus Harms, Xiulan Xie, Mohamed A. Marahiel, Xanthomonins I-III: A New Class of Lasso Peptides With a Seven-Residue Macrolactam Ring,

Angew. Chem. Int. Ed. Engl. **2014**, *53(8)*, 2230-2234.

doi: 10.1002/anie.201309267

Angew. Chem. **2014**, 126(8), 2262-2266.

doi: 10.1002/ange.201309267

Author contributions:

The project was designed by Julian D. Hegemann and Mohamed A. Marahiel with the help of Marcel Zimmermann. Experimental work was done almost exclusively by Julian D. Hegemann. The exception to this were the NMR based structure elucidation of xanthomonin II, performed by Xiulan Xie, and the cloning of the xanthomonin III gene cluster as well as the isolation of the corresponding lasso peptide, which was carried out by Shaozhou Zhu. The crystallization and crystal structure elucidation of xanthomonin I was done in a collaboration of Julian D. Hegemann, Klaus Harms and Holger Steuber. The remaining data was evaluated and interpreted by Julian D. Hegemann with assistance of Marcel Zimmermann. The manuscript was written by Julian D. Hegemann and revised after consultation with Mohamed A. Marahiel and Marcel Zimmermann. Mohamed A. Marahiel was also responsible for the overall project supervision.

Background and Summary:

Since the first lasso peptides were isolated and since their intriguing and unique three-dimensional structures were recognized, only lasso peptides with eight- or nine-residue macrolactam rings were discovered.[10,17,18,22-25,27-30,32,34,37,40,41,43-46,48,126] Consequently, it was proposed that lasso peptides could be limited to these ring sizes. In this regard, it was assumed that a ten-residue ring would be too big to allow the steric maintenance of the lariat knot, while it was doubted that the inside of a seven-residue ring would be big enough to allow the threading by the *C*-terminal tail. At least the latter assumption was proven wrong by this study, where three novel lasso peptides, xanthomonins I-III, were reported that for first time combined the lasso fold with a seven-residue macrolactam ring. Subsequent structure elucidations of xanthomonin I by X-ray crystallography and xanthomonin II by NMR spectroscopy indeed showed how small the opening in such a ring truly is. Complementary mutational studies additionally confirmed that the residues threading the ring are limited to glycine. At the same time, this small opening allowed that the lasso fold in these scaffolds can be maintained in a heat stable manner by plug amino acids as small as threonine or valine. Indeed, every amino acid bigger than serine, even when located in some distance to the position of the ring-threading, was shown to be able to maintain a thermally stable lasso fold in xanthomonin II. A comparison to the chemically accessible class of [2]rotaxanes, whose most compact representatives have a threaded ring consisting of merely 20 atoms,[127-129] shows that the 23 atom comprising macrolactam rings of the xanthomonins are only slightly bigger. This observation further emphasizes that these lasso peptides are close to the limit of what appears to be chemically possible, which makes the existence of lasso peptides with six-residue macrolactam rings rather unlikely. Still, the question for the possibility of both lasso peptides with six-residue and ten-residue macrolactam rings can only be conclusively answered by future studies.

Lasso Peptides

DOI: 10.1002/anie.201309267

Xanthomonins I–III: A New Class of Lasso Peptides with a Seven-Residue Macrolactam Ring**

*Julian D. Hegemann, Marcel Zimmermann, Shaozhou Zhu, Holger Steuber, Klaus Harms, Xiulan Xie, and Mohamed A. Marahiel**

Abstract: *Lasso peptides belong to the class of ribosomally synthesized and post-translationally modified peptides. Their common distinguishing feature is an N-terminal macrolactam ring that is threaded by the C-terminal tail. This lasso fold is maintained through steric interactions. The isolation and characterization of xanthomonins I–III, the first lasso peptides featuring macrolactam rings consisting of only seven amino acids, is now presented. The crystal structure of xanthomonin I and the NMR structure of xanthomonin II were also determined. A total of 25 variants of xanthomonin II were generated to probe different aspects of the biosynthesis, stability, and fold maintenance. These mutational studies reveal the limits such a small ring imposes on the threading and show that every plug amino acid larger than serine is able to maintain a heat-stable lasso fold in the xanthomonin II scaffold.*

Lasso peptides are a fascinating group of natural compounds that contain the unique structural motif of a lariat knot. This structural feature consists of a macrolactam ring, formed between an N-terminal α-amino group and the carboxylic acid side chain of an Asp or Glu residue, through which the C-terminal tail is threaded and interlocked through amino acids with sterically demanding side chains. These amino acids, the so-called plugs, are positioned above and below the ring.[1–16] As a consequence of their structure, lasso peptides often exhibit a high stability against proteolytic and chemical degradation, and in some cases also show exceptional thermal stability.[1,3,8,10,14,16–23] They are of general interest not only because of their unprecedented fold and conservation of their noncovalent shape, but also because they are associated with interesting biological properties, which range from antimicrobial[3,10,11,18,20,24,25] to receptor antagonistic[26–28] and inhibitory activities.[1,20,24,25,29–32] As their biosynthetic machinery is highly promiscuous, most residues of their scaffolds can be exchanged, which allowed their successful application for epitope grafting.[32]

Recently, an increasing number of new lasso peptides, particularly of proteobacterial origin, were identified.[2,14,15,23,33] Interestingly, all the lasso peptides known so far feature macrolactam rings consisting of either eight or nine amino acids. These findings gave rise to the notion of lasso peptides being limited to eight- and nine-residue rings, with ten-residue rings being proposed to be too large and seven-residue rings to be too small to maintain the lasso fold.

However, in an extensive genome-mining study on proteobacteria,[23] a small subgroup of putative precursors was found that carried a Glu residue suitable for ring formation at the seventh position of the proposed lasso peptide sequence. This was observed for 7 out of 74 of the suitable precursors identified. To probe if lasso peptides with a seven-residue ring could exist, we chose the putative biosynthetic gene cluster from *Xanthomonas gardneri* for a thorough investigation (Figure 1).

For heterologous expression, the complete cluster was cloned into pET41a. After fermentation in M9 minimal medium at 20°C for 3 days, production of the four amino acid truncated lasso peptide derived from XgaA1 and of the six amino acid truncated lasso peptide derived from XgaA2 could be detected by high-resolution FT mass spectrometry. These compounds were named xanthomonin I and II, respectively. The M9 medium was chosen for expression as it had previously been shown to be the most suitable medium for the heterologous production of lasso peptides.[14,16,23]

As the yield of this initial system was very low, production optimization was carried out as previously described.[14] Compared to the unmodified expression system, the single-precursor constructs (Figure 1c and d) showed a 37- or 11-fold increase in the overall production. By using these constructs, we were able to isolate the four amino acid truncated xanthomonin I in a yield of 12.7 $mg\,L^{-1}$ culture and the six amino acid truncated xanthomonin II in a yield of 4.1 $mg\,L^{-1}$ culture.

Analysis of the primary structure was done by MS^2 and its result was in accordance with the predicted amino acid sequences and the presence of seven-residue rings. To confirm the ring size and lasso topology, both compounds were subjected to crystallization screens, which resulted in the formation of crystals in both cases. While the crystals of the truncated xanthomonin I had a sharp outline and a defined structure, the truncated xanthomonin II formed irregularly shaped crystalline plates. As expected, the latter crystals diffracted poorly, whereas the former crystals yielded high-resolution (0.8 Å) diffraction data that allowed the elucida-

[*] J. D. Hegemann, M. Zimmermann, S. Zhu, Dr. H. Steuber, Dr. K. Harms, Dr. X. Xie, Prof. Dr. M. A. Marahiel
Fachbereich Chemie
Fachgebiet Biochemie und LOEWE-Zentrum für Synthetische Mikrobiologie, Philipps-Universität Marburg
Meerwein-Strasse 4, 35032 Marburg (Germany)
E-mail: marahiel@staff.uni-marburg.de

[**] Financial support from the Deutsche Forschungsgemeinschaft, the LOEWE program of the State of Hesse, and the Helmholtz-Zentrum Berlin is gratefully acknowledged. We would also like to thank Elena Schütz and Lara Golde for initial help with these studies, and the staff of the Bessy II MX department for providing beamtime, equipment, and support for data collection.

Supporting information for this article is available on the WWW under http://dx.doi.org/10.1002/anie.201309267.

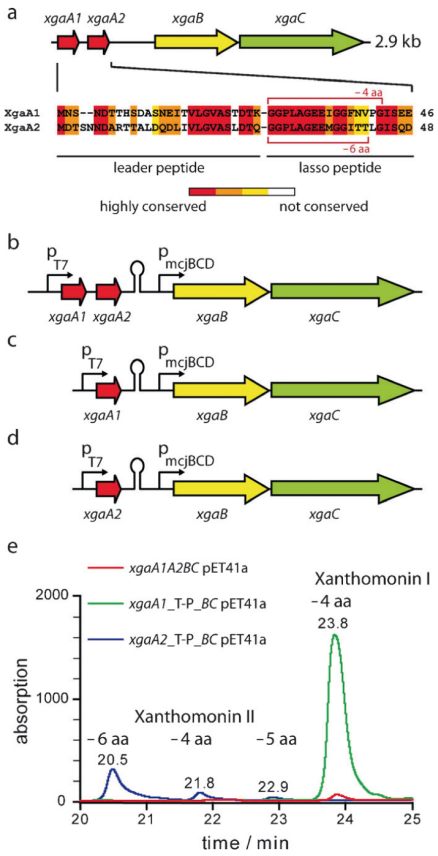


Figure 1. a) Schematic representation of the biosynthetic gene cluster from *X. gardneri* and alignment of the amino acid sequences of the identified precursor peptides. Genes were named according to the established nomenclature. The major product for each precursor is marked. b) Gene cluster after exchange of the intergenic region between *xgaA2* and *xgaB* with the λt_0 terminator (represented as a loop) and *mcjBCD* promoter sequences. c, d) Single-precursor constructs. e) UV traces of the fermentation extracts of the different constructs.

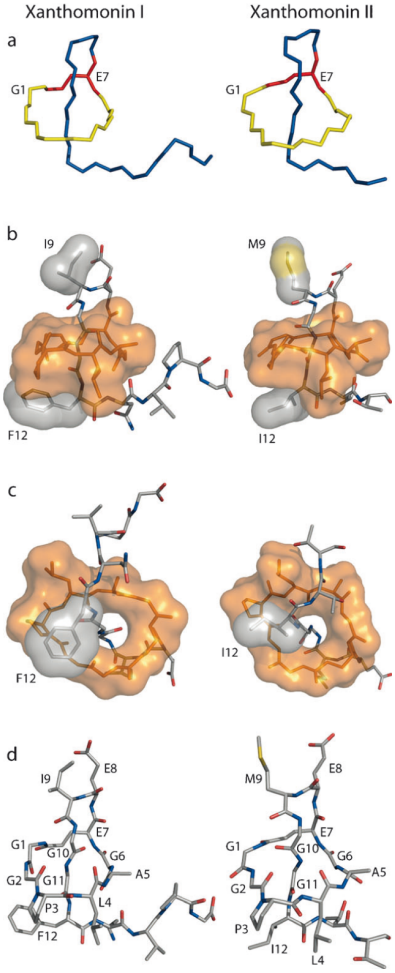


Figure 2. Several representations of the structures of the four amino acid truncated xanthomonin I (left side, PDB code 4NAG) and the six amino acid truncated xanthomonin II (right side, PDB code 2MFV). a) Structures of the lasso peptide backbones. The tail is highlighted in blue, the ring in yellow, and the ring-forming Glu7 residue in red. b, c) The interactions between the C-terminal tail and the macrolactam rings seen from the side (b) and from below (c). d) Structures of the lasso peptides with their side chains colored by elements (nitrogen in blue, oxygen in red, sulfur in yellow, carbon in gray).

tion of the structure of the four amino acid truncated xanthomonin I. This is to date only the second example of a crystal structure of a lasso peptide.[13] The structure of the six amino acid truncated xanthomonin II was solved by using NMR spectroscopic methods. The two structures are shown in Figure 2 (for more details see the Supporting Information). The plug amino acid pairs could be identified as I9/F12 for xanthomonin I and M9/I12 for xanthomonin II.

Several assays were performed to test the physicochemical properties of the xanthomonins (see the Supporting Information). The proteolytic stability was assessed against carboxypeptidase Y, chymotrypsin, and proteinase K. In almost all cases the lasso peptides proved to be completely resistant to proteolytic degradation. The only exception was that carboxypeptidase Y was able to slowly cleave off the two C-terminal amino acids of the already four amino acid truncated xanthomonin I to yield detectable traces of peptides truncated by five and six amino acids. To test the thermal stability both compounds were incubated for up to 8 h at 95 °C. In both cases, no alterations in the retention time occurred, which shows the high thermal stability of the xanthomonins.

These novel lasso peptides were also assessed for their potential antimicrobial activity through spot-on-lawn assays, but showed no activity against any of the tested bacterial strains (see the Supporting Information).

To probe the specificity of the biosynthetic machinery and to further investigate important structural features, a total of 25 xanthomonin II mutants were generated by site-directed ligase-independent mutagenesis.[34,35] Xanthomonin II was chosen for these experiments as the shorter main product facilitates the investigation of the tail region below the ring. An overview of the production of all the variants is shown in Figure 3.

Exchange of the Thr residue at the penultimate position of the leader peptide revealed that the xanthomonin system shows much lower tolerance to substitutions at this position

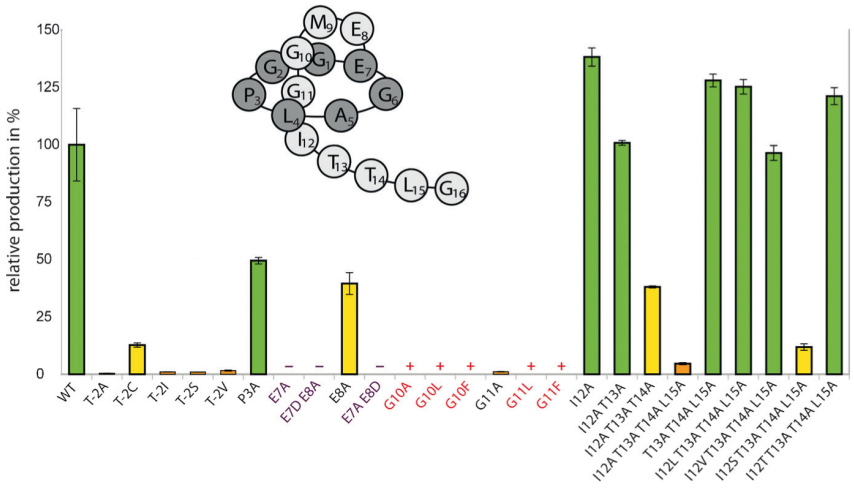


Figure 3. Overview of the overall production of all the tested xanthomonin II variants (see also the Supporting Information). A green bar represents a production of 150–50%, a yellow bar a production of 50–10%, and an orange bar a production of 10–1% compared to the wild type. Variants marked with a red cross were only detectable by mass spectrometry, while variants marked with a violet minus sign were not detectable. A schematic representation of the four amino acid truncated xanthomonin II is shown in the middle.

than other systems. Amino acids structurally similar to Thr are usually tolerated rather well, while others dramatically decrease the overall production of the lasso peptides.[14,16,22,36,37] In the case of the xanthomonin system, not only the T-2A variant, which should serve as a negative control, but also the variants T-2I, T-2S, and T-2V resulted in poor yields. Only the T-2C substitution was tolerated to some extent.

Several mutants were generated to investigate if differently sized macrolactam rings could be generated by either substituting Glu with Asp or replacing Glu7 with Ala. Furthermore, we wanted to assess if Pro3 is crucial for maturation, as the rigidity of Pro3 could be important for the prefolding. While the variants P3A and E8A were produced in good amounts, the mutations E7A, E7D, and E7A/E8D completely abolished production, thus highlighting that the biosynthetic machinery appears to be limited to processing 23-atom macrocycles.

Both amino acids that thread through the ring are glycine residues, which gives rise to the question, if this is the lower limit for a seven-residue ring. To answer this, both G10 and G11 were substituted independently with Ala, Leu, or Phe. The results clearly show that the exchange of either Gly residue with any larger amino acid almost completely abolishes production of the lasso peptide.

Finally, we wanted to investigate how small an amino acid can be to still act as a plug for a seven-residue ring. For this, we first generated a couple of variants to test if and which amino acids of the tail, besides I12, are still capable of maintaining the lasso fold. The variants I12A, I12A/T13A, and I12A/T13A/T14A were produced in good amounts. In contrast, the quadruple Ala substitution (I12A/T13A/T14A/L15A) significantly decreased the production level. Additionally, the main product of this variant was the peptide truncated by 10 amino acids, which corresponds to the macrolactam ring with only three further amino acids as a tail, which is too short for formation of a lariat knot. Based on this data, it is likely that these mutations still allow processing of the precursor, but it also suggests that the maturation is directly followed by unthreading and subsequent proteolytic degradation of the majority of the tail region.

Hence, this variant is the ideal starting point for an investigation of the role of the amino acid plug. A set of I12X/T13A/T14A/L15A variants was thus generated, with X being Ile, Leu, Ser, Thr, or Val. From these, only the variant with Ser was produced in rather poor amounts, as well as having a peptide truncated by 10 amino acids as one of its main products. To assess the topology and heat stability of these variants, a combined carboxypeptidase Y and thermal denaturation assay was performed (see the Supporting Information). While the variants carrying Ile, Leu, Thr, or Val at position 12 could not be shortened by more than six amino acids, which is in accordance with the results of the control reaction, both variants with Ala or Ser at position 12 were degraded until only the macrolactam ring and Glu8 were left.

These results were obtained independently from prior incubation at 95 °C for 4 h, thus showing that no thermally induced changes of the topology occur. Consequently, all compounds that carry Ile, Leu, Thr, or Val at position 12 are heat-stable lasso peptides, while the variants with Ala or Ser at position 12 could only be isolated as branched cyclic peptides. This shows that all amino acids larger than Ser are capable of maintaining the lasso fold of a seven-residue ring, even at high temperatures. This is a remarkable result, as similar studies with caulosegnin I revealed that Leu is barely able to maintain the lasso peptide with an eight-residue ring.[14]

It is also intriguing to see how small the openings in the macrolactam rings of the xanthomonins truly are. The fact that their rings consist of only 23 atoms allows a comparison to the class of chemically accessible [2]rotaxanes, which show some basic structural similarities to lasso peptides. Interestingly, the smallest currently known [2]rotaxanes have macrocycles consisting of only 20 atoms and, thus, are only slightly smaller than the xanthomonin rings.[38–40] Nevertheless, in contrast to a peptide chain, the macrocycles and the groups threading through them consist of only linear chains of atoms in the [2]rotaxanes. Considering this and the fact that these [2]rotaxanes are believed to be the limit of what is possible, or at least to be close to it, it is doubtful that lasso peptides with a six-residue ring could exist. However, a lasso peptide with a seven-residue ring and an Asp instead of a Glu residue as the ring-forming amino acid could still be feasible, even though the xanthomonin system was unable to process the corresponding variant. Unfortunately, there are so far no

known gene clusters with putative precursors that feature an Asp7 residue; therefore, the answer to the question, if such a lasso peptide could be naturally occurring, has to be postponed until further genomic data are available.

To prove that the reported seven-ring lasso peptides were not a special case, we decided to further investigate the gene cluster from *Xanthomonas citri* pv. *mangiferaeindicae* (see the Supporting Information), which was also identified in a previous study.[23] Production optimization was carried out as described before,[14,23] thereby increasing the overall yield fourfold and allowing the isolation of the predicted compound, xanthomonin III, truncated by seven amino acids in a yield of 0.3 mg_L^{-1} culture. MS^2 confirmed both the predicted amino acid sequence and the seven-residue ring. Combined thermal and carboxypeptidase Y studies resulted in no alteration of the length or retention time of the isolated compound, strongly indicating that it is indeed a true lasso peptide.[14,23] For identification of its plug amino acid, a set of variants was generated that successively replaced all the larger amino acids in the tail from positions 14 to 11 with Ala. As the M11A/S12A/P13A/I14A variant was the only one not produced, it can be concluded that Met11 acts as a plug for xanthomonin III. These experiments show that lasso peptides with seven-residue rings can generally be synthesized and that this class of lasso peptides is not restricted to *X. gardneri*.

In conclusion, we were able to heterologously produce and isolate three novel lasso peptides from two different biosynthetic gene clusters. These lasso peptides are the first representatives of this class of natural compounds to feature a seven-residue macrolactam ring. Thus, the old paradigm that lasso peptides can only be produced with eight- and nine-residue rings could be proven wrong. Moreover, our mutational analysis highlights the limits such a small ring imposes on the lasso fold in terms of the amino acids that can be present in the inside of a seven-residue ring and at the same time reveals the minimum size for a plug amino acid that can still maintain a heat-stable lasso fold. Additionally, we presented the NMR structure of the six amino acid truncated xanthomin II and the crystal structure of the four amino acid truncated xanthomonin I, which was solved at 0.8 Å resolution.

Received: October 23, 2013
Published online: January 20, 2014

Keywords: biosynthesis · lasso peptides · macrocycles · natural products · steric hindrance

[1] D. Frechet, J. D. Guitton, F. Herman, D. Faucher, G. Helynck, B. Monegier du Sorbier, J. P. Ridoux, E. James-Surcouf, M. Vuilhorgne, *Biochemistry* **1994**, *33*, 42 – 50.
[2] S. Um, Y. J. Kim, H. Kwon, H. Wen, S. H. Kim, H. C. Kwon, S. Park, J. Shin, D. C. Oh, *J. Nat. Prod.* **2013**, *76*, 873 – 879.
[3] O. Potterat, H. Stephan, J. W. Metzger, V. Gnau, H. Zähner, G. Jung, *Liebigs Ann. Chem.* **1994**, 741 – 743.
[4] K. L. Constantine et al., *J. Biomol. NMR* **1995**, *5*, 271 – 286.
[5] D. J. Detlefsen et al., *J. Antibiot.* **1995**, *48*, 1515 – 1517.
[6] R. Katahira, K. Shibata, M. Yamasaki, Y. Matsuda, M. Yoshida, *Bioorg. Med. Chem.* **1995**, *3*, 1273 – 1280.
[7] M. J. Bayro, J. Mukhopadhyay, G. V. Swapna, J. Y. Huang, L. C. Ma, E. Sineva, P. E. Dawson, G. T. Montelione, R. H. Ebright, *J. Am. Chem. Soc.* **2003**, *125*, 12382 – 12383.
[8] K. J. Rosengren, R. J. Clark, N. L. Daly, U. Goransson, A. Jones, D. J. Craik, *J. Am. Chem. Soc.* **2003**, *125*, 12464 – 12474.
[9] K. A. Wilson, M. Kalkum, J. Ottesen, J. Yuzenkova, B. T. Chait, R. Landick, T. Muir, K. Severinov, S. A. Darst, *J. Am. Chem. Soc.* **2003**, *125*, 12475 – 12483.
[10] M. Iwatsuki, H. Tomoda, R. Uchida, H. Gouda, S. Hirono, S. Omura, *J. Am. Chem. Soc.* **2006**, *128*, 7486 – 7491.
[11] T. A. Knappe, U. Linne, S. Zirah, S. Rebuffat, X. Xie, M. A. Marahiel, *J. Am. Chem. Soc.* **2008**, *130*, 11446 – 11454.
[12] T. A. Knappe, U. Linne, X. Xie, M. A. Marahiel, *FEBS Lett.* **2010**, *584*, 785 – 789.
[13] H. Nar, A. Schmid, C. Puder, O. Potterat, *ChemMedChem* **2010**, *5*, 1689 – 1692.
[14] J. D. Hegemann, M. Zimmermann, X. Xie, M. A. Marahiel, *J. Am. Chem. Soc.* **2013**, *135*, 210 – 222.
[15] M. O. Maksimov, A. J. Link, *J. Am. Chem. Soc.* **2013**, *135*, 12038 – 12047.
[16] M. Zimmermann, J. D. Hegemann, X. Xie, M. A. Marahiel, *Chem. Biol.* **2013**, *20*, 558 – 569.
[17] D. F. Wyss, H. W. Lahm, M. Manneberg, A. M. Labhardt, *J. Antibiot.* **1991**, *44*, 172 – 180.
[18] R. A. Salomon, R. N. Farias, *J. Bacteriol.* **1992**, *174*, 7428 – 7435.
[19] M. Yamasaki, K. Yano, M. Yoshida, Y. Matsuda, K. Yamaguchi, *J. Antibiot.* **1994**, *47*, 276 – 280.
[20] K. Yano, S. Toki, S. Nakanishi, K. Ochiai, K. Ando, M. Yoshida, Y. Matsuda, M. Yamasaki, *Bioorg. Med. Chem.* **1996**, *4*, 115 – 120.
[21] A. Blond, M. Cheminant, D. Destoumieux-Garzón, I. Ségalas-Milazzo, J. Peduzzi, C. Goulard, S. Rebuffat, *Eur. J. Biochem.* **2002**, *269*, 6212 – 6222.
[22] T. A. Knappe, U. Linne, L. Robbel, M. A. Marahiel, *Chem. Biol.* **2009**, *16*, 1290 – 1298.
[23] J. D. Hegemann, M. Zimmermann, S. Zhu, D. Klug, M. A. Marahiel, *Biopolymers* **2013**, *100*, 527 – 542.
[24] M. Tsunakawa et al. *J. Antibiot.* **1995**, *48*, 433 – 434.
[25] K. Kimura, F. Kanou, H. Takahashi, Y. Esumi, M. Uramoto, M. Yoshihama, *J. Antibiot.* **1997**, *50*, 373 – 378.
[26] Y. Morishita, S. Chiba, E. Tsukuda, T. Tanaka, T. Ogawa, M. Yamasaki, M. Yoshida, I. Kawamoto, Y. Matsuda, *J. Antibiot.* **1994**, *47*, 269 – 275.
[27] T. Ogawa, K. Ochiai, T. Tanaka, E. Tsukuda, S. Chiba, K. Yano, M. Yamasaki, M. Yoshida, Y. Matsuda, *J. Antibiot.* **1995**, *48*, 1213 – 1220.
[28] O. Potterat, K. Wagner, G. Gemmecker, J. Mack, C. Puder, R. Vettermann, R. Streicher, *J. Nat. Prod.* **2004**, *67*, 1528 – 1531.
[29] J. Yuzenkova, M. Delgado, S. Nechaev, D. Savalia, V. Epshtein, I. Artsimovitch, R. A. Mooney, R. Landick, R. N. Farias, R. Salomon, K. Severinov, *J. Biol. Chem.* **2002**, *277*, 50867 – 50875.
[30] K. Adelman, J. Yuzenkova, A. La Porta, N. Zenkin, J. Lee, J. T. Lis, S. Borukhov, M. D. Wang, K. Severinov, *Mol. Cell* **2004**, *14*, 753 – 762.
[31] E. Semenova, Y. Yuzenkova, J. Peduzzi, S. Rebuffat, K. Severinov, *J. Bacteriol.* **2005**, *187*, 3859 – 3863.
[32] T. A. Knappe, F. Manzenrieder, C. Mas-Moruno, U. Linne, F. Sasse, H. Kessler, X. Xie, M. A. Marahiel, *Angew. Chem.* **2011**, *123*, 8873 – 8876; *Angew. Chem. Int. Ed.* **2011**, *50*, 8714 – 8717.
[33] M. O. Maksimov, I. Pelczer, A. J. Link, *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 15223 – 15228.
[34] J. Chiu, P. E. March, R. Lee, D. Tillett, *Nucleic Acids Res.* **2004**, *32*, e174.
[35] J. Chiu, D. Tillett, I. W. Dawes, P. E. March, *J. Microbiol. Methods* **2008**, *27*, 195 – 198.
[36] S. J. Pan, J. Rajniak, M. O. Maksimov, A. J. Link, *Chem. Commun.* **2012**, *48*, 1880 – 1882.

[37] W. L. Cheung, S. J. Pan, A. J. Link, *J. Am. Chem. Soc.* **2010**, *132*, 2514 – 2515.
[38] S. Dasgupta, K. W. Huang, J. Wu, *Chem. Commun.* **2012**, *48*, 4821 – 4823.
[39] S. Dasgupta, J. Wu, *Chem. Sci.* **2012**, *3*, 425 – 432.
[40] H. W. Gibson, D. S. Nagvekar, N. Yamaguchi, S. Bhattacharjee, H. Wang, M. J. Vergne, D. M. Hercules, *Macromolecules* **2004**, *37*, 7514 – 7529.

3.5 Characterization of Caulonodin Lasso Peptides Revealed Unprecedented N-Terminal Residues and a Precursor Motif Essential for Peptide Maturation

Marcel Zimmermann, Julian D. Hegemann, Xiulan Xie, and Mohamed A. Marahiel,
Characterization of Caulonodin Lasso Peptides Revealed Unprecedented N-Terminal Residues and a Precursor Motif Essential for Peptide Maturation,
Chem. Sci. **2014**, *Epub ahead of print*.
doi: 10.1039/c4sc01428f

Author contributions:

The project was designed by Marcel Zimmermann and Mohamed A. Marahiel with the assistance of Julian D. Hegemann. The experiments were carried out by Marcel Zimmermann with the only exceptions being the NMR based structure elucidation of caulonodin V and the NMR based structural investigation of caulonodin VI, which both were performed by Xiulan Xie. Data evaluation and interpretation was done by Marcel Zimmermann with the aid of Julian D. Hegemann. The manuscript was written by Marcel Zimmermann and revised with the support of Julian D. Hegemann and Mohamed A. Marahiel, the latter of whom was also responsible for the overall project supervision.

Background and Summary:

Before this study, every known lasso peptide featured either a cysteine or a glycine residue at position 1, depending on it being either a class I or a class II/III lasso peptide.[10,17,18,22-30,32,34,37,40,41,43-46,48,126] As this observation held true since the discovery of the first lasso peptide in 1991,[46] it was assumed that the conservation of the position 1 residue was a general criteria valid for all lasso peptides. In this work, two gene clusters from *Caulobacter* sp. K31 were tested for heterologous lasso peptide production in *E. coli*. In contrast to all lasso peptide biosynthetic gene clusters that were investigated before, the four corresponding lasso peptides of these two gene clusters would carry either an alanine or serine residue at their *N*-terminus. Through this, we tried to assess if these biosynthetic systems were indeed able to produce class II lasso peptides with macrolactam rings formed between a carboxylic acid side chain and an *N*-terminal residue different than glycine. This was proven to be true for all four proposed lasso peptides, allowing us to revise another paradigm of lasso peptide research. To substantiate these findings, the NMR structure of caulonodin V was elucidated and the long-range NOE contacts in caulonodin VI were investigated in order to clarify the identity of its upper and lower plug amino acids. To prove that the plugs of caulonodins IV and VII were at the corresponding positions as the ones of their sibling compounds, a thorough mutational and stability analysis was performed. Finally, these new results were put into context of our previous genome mining study[37] and a bioinformatic analysis of all known functional lasso peptide precursors was done that led to the identification of a *C*-terminal leader peptide motif that appears to be highly conserved among ABC-type proteobacterial lasso peptide biosynthetic gene clusters. The importance of this conserved motif was proven in two ways. Firstly, a mutational study with caulonodin V confirmed the significance of different residues in this motif as any non-conserved mutations of these positions drastically reduced the overall lasso peptide production. Secondly, we noticed that the leader peptide of the previously reported rhodanodin[37] naturally featured a serine residue at position -8 instead of the highly conserved glycine of the identified motif. By exchanging this serine with a glycine residue, thereby restoring the original motif, the heterologous production of rhodanodin was increased significantly, emphasizing the importance of this motif for (proteobacterial) lasso peptide maturation.

EDGE ARTICLE

Characterization of caulonodin lasso peptides revealed unprecedented N-terminal residues and a precursor motif essential for peptide maturation†

Cite this: DOI: 10.1039/c4sc01428f

Marcel Zimmermann, Julian D. Hegemann, Xiulan Xie and Mohamed A. Marahiel*

Lasso peptides, a peculiar family of ribosomally assembled and post-translationally modified peptides (RiPPs), possess a fascinating 3D structure, which can confer rigidity and stability against chemical and thermal denaturation. Their distinctive "lariat knot" structure is accountable for their antibacterial, enzyme inhibitory and receptor antagonist activities. While the biosynthetic machinery was recently characterized, the rules concerning the formation of this unique lasso structure on the basis of their peptide sequences remain elusive. Restrictions such as the length of the peptide, the size of the ring, or the nature of the amino acids associated with the lasso fold stabilization were recently overhauled by the identification of new members of this RiPP family. In this work we demonstrate the isolation of four genome-mining-predicted lasso peptides featuring the unprecedented amino acids serine or alanine at position 1 of the core peptide. By a mutational approach we were able to predict the lasso fold for four peptides (caulonodins IV to VII). This prediction was confirmed for caulonodin V by the full elucidation of its 3D-structure *via* NMR and for caulonodin VI by the determination of long range NOE-contacts. Furthermore, the substrate specificity of the biosynthetic machinery for the atypical position 1 was probed. Additionally, utilizing the recent growth of functional lasso peptide precursor sequences we were able to identify a conserved motif in the C-terminal part of the leader peptide through bioinformatics analysis. Employing an extensive *in vivo* analysis for substitution tolerance of the biosynthetic machinery in this conserved region confirmed the significance of several residues, indicating that the predicted motif is very likely a general leader peptide recognition sequence specific for lasso peptide maturation.

Received 15th May 2014
Accepted 19th June 2014

DOI: 10.1039/c4sc01428f

www.rsc.org/chemicalscience

Introduction

The constant search for novel compounds for anti-cancer treatment, as antiviral drugs or antibacterial agents against multi-resistant strains presents one of the great challenges in the 21st century. The discovery and characterization of new natural compounds is the basis for this endeavor. These molecules can possess a wide variety of intrinsic bioactivities or may be useful as scaffolds for drug design and epitope grafting.[1,2] While selective screenings for certain bioactivities have been performed in the past by bacterial or fungal fermentations, nowadays the rational approaches gain more ground.[3,4] In particular, genome mining driven predictions and isolations of new compound become more common[5,6] as sequencing costs drop dramatically. Thousands of bacterial genomes have been already sequenced, providing large amounts of data to be analyzed in upcoming research.

A vast group of natural compounds that are readily accessible to genome mining approaches are the ribosomally assembled and post-translationally modified peptides (RiPPs).[5,7] The gene clusters responsible for the production of these compounds contain several processing enzymes that transform a precursor peptide, consisting of a leader and a core peptide into the mature product.[8] Therefore, their gene clusters are often smaller than those encoding non-ribosomal peptide synthetases (NRPS) or poly ketide synthases (PKS) with products of comparable complexity. There are nearly two dozen families of RiPPs classified by their chemical modifications and biochemical routes to these modifications.[7] One family of RiPPs that is rather simple from a chemical point of view, but possesses an intriguing 3D-structure as its distinctive feature, are the lasso peptides.[9]

Department of Chemistry, Biochemistry, Philipps-University Marburg, Hans-Meerwein-Strasse 4 and LOEWE-Center for Synthetic Microbiology, D-35032, Marburg, Germany. E-mail: marahiel@staff.uni-marburg.de

† Electronic supplementary information (ESI) available: Heat and protease stability assays, 3D fold schemes, ^{1}H-NMR and 2D NOE spectra of caulonodin V, Ramachandran plot, ^{1}H-NMR and 2D NOE spectra of caulonodin VI, caulonodin precursor alignment, MEME leader alignment, a list of all used oligonucleotide primers, contents of M9 vitamin mix, overview of possible plug amino acids, detailed production amounts of all created mutants, ^{1}H chemical shift assignment and structural statistics of caulonodin V and a partial ^{1}H chemical shift assignment of caulonodin VI. See DOI: 10.1039/c4sc01428f

These RiPPs with a knotted structure[10–12] are produced from an approx. 50 amino acid (aa) precursor, peptide A, by a reaction mechanism involving two interdependent enzymes.[13] The first enzyme B was shown to be a cysteine protease, while protein C shows cyclase activity.[14] The latter reaction is mediated *via* an AMP-activation of an acidic side chain, which then attacked by the newly generated N-terminus of the core peptide leads to macrolactam ring formation. The C-terminal part of the peptide is in the meantime threaded through the ring in a presumable prefolding reaction that precedes the ring closure.[8] For both reactions catalyzed by the B and the C protein, a complex formation as well as ATP hydrolysis is essential.[13] The energy consumption of protein B suggests its involvement in precursor binding and prefolding. After the assembly, export might occur by an ATP-binding cassette (ABC) transporter encoded from the lasso gene cluster (as for microcin J25 and capistruin) or by another endogenous transport system, conferring immunity to the producer strain.[15–17] The functions of lasso peptides in the producer strains remain elusive, although the recent discovery of a specific isopeptidase gave rise to the idea that lasso peptides may act as scavenging or signaling molecules.[18] This family of RiPPs can be further divided into classes according to the presence (classes I and III) or absence (class II) of conserved cysteine residues involved in the formation of up to two disulfide bonds.[19,20]

Since the discovery of the first lasso peptides anantin[21,22] and microcin J25 (ref. 23) in the early 1990s and the first reported lasso structure[24] in 1994, several paradigms were established for this class of RiPPs. These include the extraordinary stability against temperature, the necessity for bulky side chain amino acids in the C-terminal tail, the restriction of ring size to eight or nine amino acids and the presence of a glycine or cysteine at the first position of the core peptide.[25,26] Three of these doctrines were recently overhauled. Heat sensitive lasso peptides were recently discovered[27] and it was shown that small amino acids are capable of entrapping the tail under certain circumstances.[28] Furthermore, three lasso peptides with a ring size of only seven amino acids have just been reported.[29] While the paradigm of the first amino acid to be Gly or Cys was still sustained, several mutagenesis approaches have shown that although the tolerance for substitution at this initial position is highly restricted,[30,31] the biosynthetic machinery of the caulosegnins is capable of processing detectable amounts of G1X variants.[28] To investigate the possibility of naturally occurring non-Gly1 class II lasso peptides, clusters from a recent expanded genome mining approach in proteobacteria were inspected more closely.[26] These clusters have a simple ABC gene organization but feature precursors that do not fit the criterion of Gly at position 1 of the core peptide. Two of these clusters, which have been previously identified as potential lasso peptide biosynthetic gene clusters,[32] are present in the genome of *Caulobacter* sp. K31 In total, this strain carries three lasso peptide biosynthetic gene clusters, from which only one, produces three regular Gly1 type lasso peptides,[26] whereas the other clusters suggest different N-terminal residues.

Materials and methods

Strains and general methods

Caulobacter sp. K31 was purchased from the German Collection for Microorganisms and Cell Cultures (DMSZ). *E. coli* TOP10, which was used for cloning, and *E. coli* BL21(DE3), which was used for heterologous expression, were purchased from Invitrogen. Oligonucleotides and carboxypeptidase Y were purchased from Sigma Aldrich. Trypsin was purchased from Promega. Restriction enzymes, Phusion polymerase and T4 DNA Ligase were purchased from New England Biolabs. The identity of constructed plasmids and mutants was determined with DNA Sanger Sequencing and was performed by GATC Biotech AG (Konstanz).

Cloning of the caulonodin biosynthesis clusters

Genomic DNA of *Caulobacter* sp. K31 was used to amplify the *Caul19xxAABC* and the *Caul22xxAABC* gene cluster with PCR (for primers see ESI Table S1†). PCR was performed with Phusion Polymerase following the instructions of the manufacturer. PCR was subjected to gel electrophoresis and amplicons with the correct size were excised, purified and subsequently digested with NdeI and XhoI. pET41a vector was digested in the same way and the fragments were ligated with T4 DNA Ligase. Transformation was performed with TOP10 cells for cloning and BL21(DE3) for expression.

Engineering of the clusters and construction of caulonodin and rhodanodin mutants

The *Caul19xxAABC*pET41a plasmid and the *Caul22xxAABC*-pET41a plasmid were used to engineer the clusters by deletion of the intergenic region between the second precursor and the gene encoding the B-protein and incorporation of a ribosomal binding site. Both these plasmids were further modified to create 4 single precursor plasmids by deletion of one of the precursors: *Caul1984A1rbsBC*pET41a, *Caul198xA2rbsBC*pET41a, *Caul2238A1rbsBC*pET41a, *Caul2239A2rbsBC*pET41a. The cluster engineering was entirely done with site-directed-ligation-independent-mutagenesis (SLIM).[33]

These single precursor constructs were used for mutagenesis to create lasso peptide variants and for the investigation of the lasso precursor. These follow up mutagenic modifications were either done with a modified protocol of the site directed mutagenesis with inverse PCR[34] or with the SLIM protocol. The *rhotA_RBS_BC* pET41a plasmid[26] was used to create the rhodanodin variant. Primers for each mutant or mutant group are shown in ESI Table S1.†

Heterologous expression of the WT and mutated caulonodin and rhodanodin constructs

E. coli BL21(DE3) cells were transformed with one of the caulonodin or rhodanodin plasmids or mutants thereof and grown overnight at 37 °C in LB medium containing kanamycin (50 μg mL^{-1}). M9 minimal medium supplemented with M9 vitamin mix (ESI Table S2†) and containing kanamycin (50 μg mL^{-1}) was

inoculated with the overnight culture to an OD_{600} of 0.01. The culture was grown at 37 °C to an OD_{600} of 0.45. Then the temperature was shifted to 20 °C, the cells were grown for 45 min to adapt to the new growth conditions and the expression was induced with the addition of IPTG to a final concentration of 0.1 mM. Expression was continued for 3 days and the cells were harvested by centrifugation. Pellets were extracted as stated in the purification section and extracts were subsequently applied to high-performance liquid chromatography (HPLC) for purification or analyzed by a high-performance liquid chromatography mass spectrometry (HPLC-MS) system.

Purification of the caulonodins IV to VII

The cell pellet (6000 rpm, 20 min) of a single precursor expression culture was extracted with methanol for 1 h at RT. The obtained solvent after extraction was evaporated to dryness with a rotational evaporator at 40 °C or in case of the variants removed *via* lyophilization.

Crude extracts of the caulonodin fermentations were resuspended in 20% acetonitrile and subsequently applied to a preparative reverse phase (RP) HPLC system (1100 series Agilent) using a C18HTec column (250 × 21 mm) with a gradient of water/0.1% trifluoroacetic acid (solvent A), acetonitrile/0.1% trifluoroacetic acid (solvent B) with a flow rate of 18 mL min^{-1}. The gradient for all four caulonodins was as follows: a linear increase from 10% B to 60% B in 30 min followed by the washing increasing to 95% B in 2 min and holding 95% B for 5 min. Retention times (R_t) of the produced caulonodins IV to VII were 19.4, 16.0, 16.1 and 18.5 min respectively. The yield of caulonodin IV was approx. 3 mg L^{-1} culture. Caulonodin VI was produced with a yield of around 10 mg L^{-1} culture, while only about 0.7 mg L^{-1} culture of caulonodin VII were obtained.

Further purification of caulonodin V for NMR spectroscopic investigation was performed with an semi preparative scale HPLC system (1260 series Agilent Technologies) with a fraction collector using a Nucleodur C18ec column (250 × 5 mm) at a column temperature of 25 °C and a flow rate of 0.8 mL min^{-1} with the following gradient using the same solvents as before: linear increase from 25% B to 32.5% B in 15 min followed by the washing performed in the same manner as the preparative scale. Caulonodin V had a R_t of 12.0 min. The final yield of caulosegnin V was 4 mg L^{-1} culture.

Mass spectrometric analysis

Mass spectrometric analysis of the caulonodin and rhodanodin extracts and variant extracts was performed with a high-resolution LTQ-FT ultra instrument (Thermo Fisher Scientific) connected to a micropore 1100 HPLC system (Agilent) using sample amounts of up to 100 μL. UV absorption spectra were recorded at 215 nm. Separation was achieved using a CC Nucleosil 300-8 C18 column (125 × 2 mm) (Macherey-Nagel) applying the following gradient of water/0.1% trifluoroacetic acid (solvent A) and acetonitrile/0.1% trifluoroacetic acid (solvent B) at a column temperature of 40 °C and a flow rate of 0.2 mL min^{-1}: starting with a linear increase from 5% to 50% B in 30 min, a subsequent increase from 50% to 95% B in 2 min and holding 95% B for additional 5 min.

To quantify the products of a wild type or mutant fermentation, UV-peak areas were integrated and relative production was determined by comparison between mutant and wild type. Fermentations were carried out in triplicates for each mutant and the averages as well as standard deviations were calculated.

Collision-induced dissociation fragmentation studies within the linear ion trap were done using online HPLC-MS. In most cases the doubly charged ions were selected for fragmentation, as they were the dominant species in the spectra. The energy for fragmentation was set to 35 for every measurement performed.

Mass spectrometric analysis after the thermal stability and protease assays of the purified caulonodins was performed with a low-resolution 1100 series MSD (Hewlett-Packard) coupled with a micropore 1260 HPLC system (Agilent Technologies).

For each of the caulonodins an adapted gradient was used. For caulonodin V a linear increase from 24% B to 31.5% B in 15 min was applied. Caulonodin VI assays were analyzed with a gradient from 23% B to 30.5% B. For caulonodin VII the gradient was further adjusted starting from 28% B and finishing with 35.5% B. The washing steps were applied analog to the purification.

Heat stability assays of the caulonodins and variants

To investigate the thermal stability of the four new caulonodin lasso peptides, a solution of 10 μg purified lasso peptide was incubated at 95 °C for up to 4 h. Samples were cooled to 4 °C and were subsequently analyzed *via* low-resolution HPLC-MS using the individual caulonodin gradients.

Heat stability of the variants was investigated by incubating 30 μL of the respective extract at 95 °C for 1 h. The samples were cooled and analyzed *via* high-resolution HPLC-MS or further treated with carboxypeptidase Y. As ref. 30, μL of the respective untreated extract were analyzed.

Protease assays of the caulonodins and variants

Stability against proteolytic degradation of the caulonodins and the branched-cyclic variants were investigated by incubating 10 μg of the purified lasso peptide before and after heat treatment (95 °C, 1 h) with carboxypeptidase Y. The carboxypeptidase Y assays were performed with 0.5 U carboxypeptidase Y in a buffer containing 50 mM MES and 1 mM $CaCl_2$ at pH = 6.75 for 4 h or 16 h at 25 °C.

Variants were not isolated and therefore 30 μL of the pellet extract were mixed with 30 μL of a carboxypeptidase Y (1.5 U) containing buffer (50 mM MES, 1 mM $CaCl_2$, pH = 6.75) and incubated for 2 h at 25 °C. To stop the digest 50 μL of water were added and the samples were frozen and stored at −20 °C until applied to high or low-resolution HPLC-MS analysis.

NMR-spectroscopic methods

Samples for NMR measurements contained 4.0 mg of caulonodin V in 200 μL H_2O–D_2O (9 : 1) or 4.5 mg caulonodin VI in 200 μL H_2O–D_2O (9 : 1), respectively in Wilmad 3 mm tubes (Rototec Spintec). Spectra were recorded on a Bruker Avance 600

MHz spectrometer equipped with an inverse triple resonance $^{1}H–^{13}C–^{15}N$ probe with z-gradient. The effect of temperature on the structures was surveyed by recording ^{1}H spectra at variable temperatures. Thus, all 2D spectra for the structure determination of caulonodin V were recorded at 298 K, while for caulonodin VI 288 and 300 K were chosen for DQF-COSY, TOCSY and NOESY spectra. For sequential assignment DQF-COSY,[35] TOCSY[36] and NOESY[37] experiments were performed in phase-sensitive mode using states-TPPI.[38] TOCSY spectrum was recorded with a mixing time of 140 ms, while NOESY spectra were observed at 150 and 300 ms mixing times. Water suppression was fulfilled using excitation sculpting[39] with gradients for DQF-COSY, TOCSY and NOESY. 1D spectra were acquired with 65 536 data points, while 2D spectra were collected using 4096 points in the F_2 dimension and 512 increments in the F_1 dimension. For 2D spectra 32 transients were used, with relaxation delay of 3.0 s. Chemical shifts were referenced to H_2O signal, which was calibrated in turn using 2,2-dimethyl-2silapentane-5-sulfonat (DSS) as internal standard in a different sample at the same temperature. All spectra were processed with Bruker TOPSPIN 3.1. NOE cross-peaks were analyzed within the program Sparky.[40]

Structure calculations for caulonodin V were performed with the program CYANA 2.1.[41] The internal linkage was realized by setting the distance constraints between N of Gly1 and Cδ of Glu9 to be 1.33 Å. NOE cross-peaks observed in the 150 ms mixing time NOESY experiment were converted into distance constraints manually. In this way, 144 unambiguous distance constraints were obtained, 48 for the backbone, 22 for long-range interactions, and 74 for the side-chains. Thus, there was an average of 8.0 distance constraints per residue. In addition, constraints of torsion angles ϕ and χ^1 were determined by analyzing the vicinal coupling constants $^{3}J_{HN\alpha}$ and $^{3}J_{\alpha\alpha}$. Only unambiguous coupling constants were used. Thus, $^{3}J_{HN\alpha} \geq 9$ Hz was observed for Ser1, Ile2, Asp4, Gly6, Thr15 and Tyr16. The torsion angles ϕ of these residues were restrained to $-120^\circ \pm 30^\circ$. For the residues Gly3, Ser5, Leu7–Gln14 and Trp17 $^{3}J_{HN\alpha} < 9$ Hz was detected. Thus their torsion angles ϕ were restrained to $-70^\circ \pm 30^\circ$. Stereospecific assignment of the following prochiral β-methylene protons were fulfilled by measuring $^{3}J_{\alpha\beta}$ and analyzing patterns of the intraresidual NOE interactions $d_{\alpha\beta}$ and $d_{N\beta}$:[42] Ser1, Ile2 and Ser10 in (g^2g^3); Glu9, Tyr16 and Trp17 in (t^2g^3). For g^2g^3 and t^2g^3 conformations around the Cα–Cβ bond the torsion angle χ^1 was constrained in the range of $-60 \pm 30^\circ$and $60 \pm 30^\circ$ respectively.

The above mentioned constraints were used in the simulated annealing protocol for calculation in CYANA 2.1 program. The calculation initiated with 50 random conformers and the resultant structures were engineered by the program package Sybyl 7.3 (ref. 43) to include the covalent linkage between the nitrogen of Ser1 and Cδ of Glu9, followed by energy minimization under NMR constraints using TRIPOS force field within Sybyl. Thus, on the basis of low energies and minimal violations of the experimental data, a family of 15 structures was chosen. These 15 energy-minimized conformers show an average root-mean-square deviation of 0.03 Å and are kept to represent the solution structure of caulonodin V (PDB accession code 2mlj).

Results and discussion

Caulobacter K31 as a rich source of lasso peptides

Previous genome mining studies[26,32,44] identified three potential lasso peptide biosynthetic gene clusters in *Caulobacter* sp. K31 (Fig. 1a). These feature a plain ABC gene organization and therefore are lacking a transport protein, while in close proximity several genes are located that are typical for proteobacterial lasso clusters, namely a GntR type regulator upstream and a TonB dependent receptor isopeptidase cluster downstream. All three clusters contain at least two genes encoding precursor peptides, raising the number of putative class II lasso peptide products derived from this strain to seven. The cluster 519x AAABC features three precursor peptides and has already been studied for production of lasso peptides in a recent B-protein-based genome mining approach focused on proteobacteria.[26] The two remaining clusters, 198x AABC and 22xx AABC, feature two precursor peptides each. While the processing enzymes share high degrees of homology to those of the characterized cluster, the precursors do not fit the general criteria defined for lasso peptides, having unusual residues at the possible cleavage site between leader and lasso sequence. For the two precursors of cluster 22xx AABC the sequence at the cleavage site is TR/AG and TR/SG, respectively. If the location of the highly conserved Thr[45] remained at the penultimate position of the leader peptide, this would lead to lasso peptides containing Ser or Ala at position 1. Applying the same assumption to the two precursors of cluster 198x AABC (cleavage site TQ/SF and TR/SI), results in two possible lasso peptides both with Ser at position 1 of the core peptide (Fig. 1b). Although more than 20 years after the discovery of the first lasso peptides[21–23] no instances have been reported where another amino acid than Gly or Cys was present at position 1 of a lasso peptide. Based on the similarity of the three lasso clusters present in *Caulobacter* sp. K31 and the fact that one of them is functional, we were encouraged to investigate the two clusters with the unprecedented N-terminal residues, Ala and Ser, for lasso peptide production with our established methods.

Heterologous production of four new lasso peptides from *Caulobacter* K31

In recent publications it was shown that the incorporation of an artificial ribosomal binding site between the genes encoding the precursor peptide and the downstream located biosynthetic genes greatly enhances the heterologous production of lasso peptides in *E. coli*.[45] Furthermore, the design of single precursor constructs was shown to be beneficial for the production and subsequent purification of lasso peptides.[28,29] Accordingly, four constructs were created for the heterologous expression of the four potential new lasso peptides from *Caulobacter* sp. K31 (Fig. 1c). Fermentations were carried out in M9 minimal medium for 3 days at 20 °C and pellet extracts were analyzed by high-resolution Fourier transform mass spectrometry system coupled to a HPLC system. In all four cases, masses were detected that correspond to the predicted full length lasso peptides, which have a serine or alanine at position 1 (Fig. 1d).

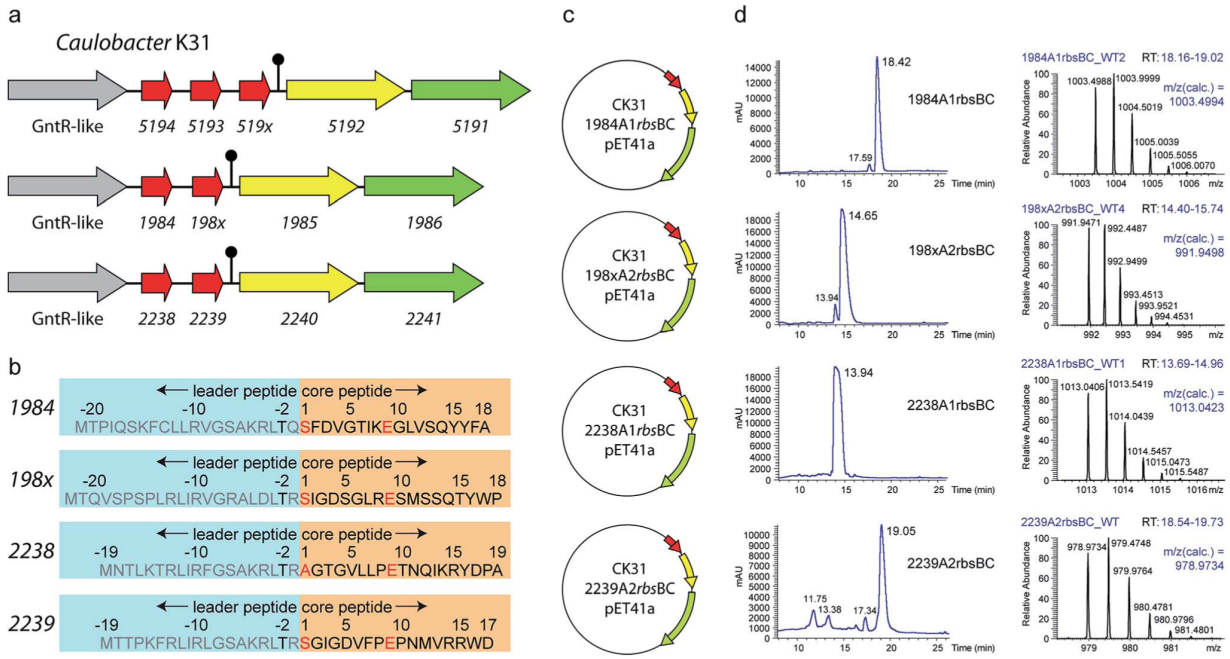


Fig. 1 (a) Overview of the three putative lasso peptide biosynthetic gene clusters, (b) the four precursors from the two uncharacterized clusters, (c) the four constructs used for heterologous expression, (d) high-resolution HPLC-MS analysis of the four constructs shown as UV traces and mass signals.

The following mass fragmentation (MS^2) analyses confirmed the identity of these peptides and in two of four cases prove the internal backbone to side chain cyclization of residue 1 (Ser or Ala) to Glu9 (Fig. 2). These four new lasso peptides, termed caulonodins IV to VII, are the first examples for native class II lasso peptides that do not start with a glycine but instead feature an alanine or serine residue at position 1.

This essentially redefines the prerequisites for class II lasso peptides, which considered glycine to be the only suitable amino acid for the N-terminus of the core peptide. Therefore, the experimental confirmation of these peptides facilitates further genome mining for lasso peptides with any amino acid at the N-terminal position. Applying this new criterion on the results of a recent genome mining study that focused on proteobacteria raises the number of suitable precursor peptides from 74 (ref. 26 and 29) to a total of 97 out of the 98 identified precursors.

Heat and protease stability of caulonodins IV to VII

Each peptide could be produced and purified from its respective single precursor construct in sufficient amounts to allow subsequent characterization. Nonetheless, the production of the four peptides significantly differed. While the fermentation of caulonodin VII yielded 0.7 mg L^{-1} culture, caulonodins IV and V were produced with approx. 3 to 4 mg L^{-1}. Caulonodin VI was produced best with a yield of 10 mg L^{-1} of culture.

All peptides were investigated for their heat stability by incubation at 95 °C and subsequent treatment with carboxypeptidase Y to distinguish lasso peptide and unfolded branched-cyclic peptide as previously described.[27]

Caulonodin IV showed an unusual behavior upon thermal denaturation. Instead of the formation of a clear second peak with an identical mass, which was observed for other heat sensitive lasso peptides, it showed a decrease of the UV signal (ESI Fig. S1a†). A closer inspection of the sample revealed the formation of precipitate, indicating a strong increase in hydrophobicity of the thermally denatured peptide lowering its solubility. Therefore, it was not possible to investigate its stability against carboxypeptidase Y.

The HPLC elution profile of caulonodin VI significantly changed after heat treatment at 95 °C. A second peak with the same mass arose as well as other peaks with masses fitting to hydrolysis products of the lasso peptide (ESI Fig. S1b†). Close monitoring of this behavior with short time intervals revealed the step by step process of unfolding and subsequent hydrolysis between Asp17 and Pro18, followed by an addition of water at an unknown position. The unfolded full length peptide as well as the truncated variants were susceptible to carboxypeptidase Y.

The Caulonodins V and VII showed no change in the chromatographic behavior even after prolonged exposure (up to 4 h) to 95 °C. Nonetheless, proteolytic digestions revealed sensitivity against carboxypeptidase Y in both cases leading to the assumption that unfolding has happened and that the chromatographic behavior of the lasso peptides and their respective branched-cyclic analogues are very similar (ESI Fig. S1c and d†).

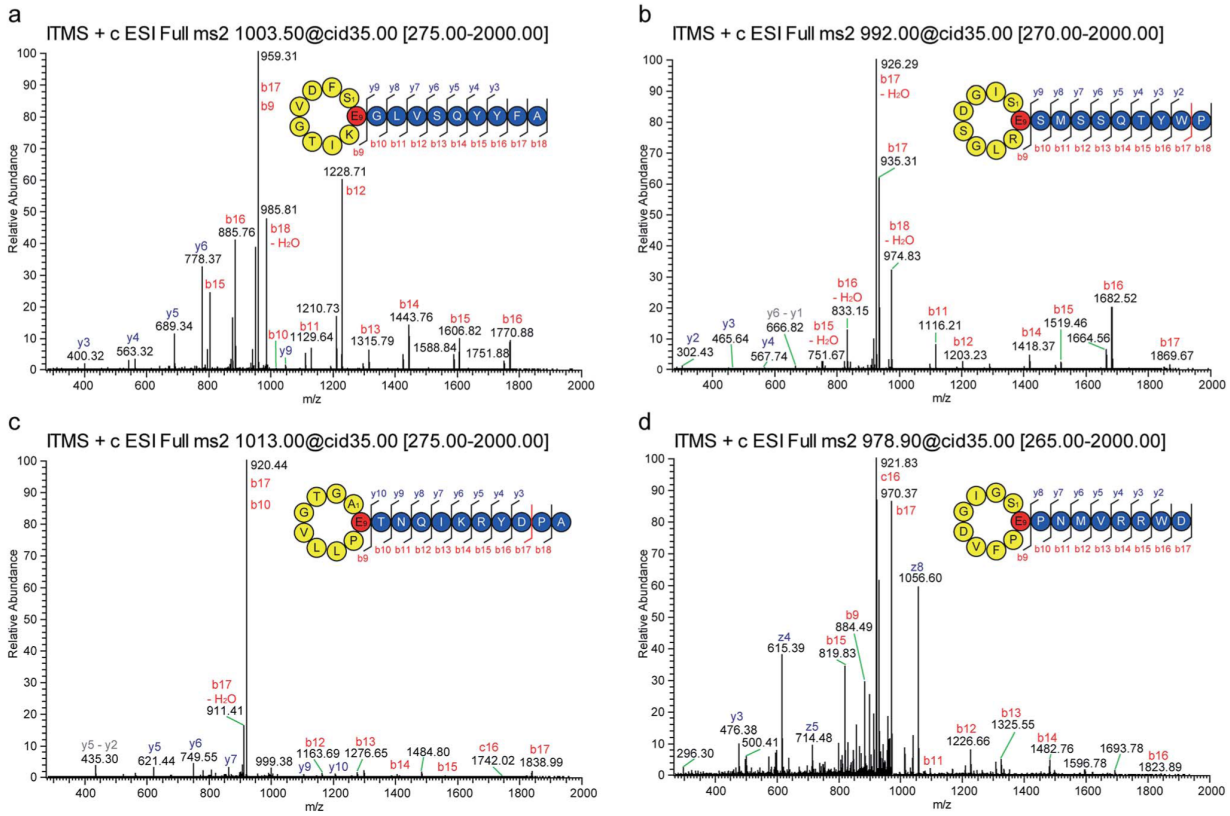


Fig. 2 MS^2 spectra of (a) caulonodin IV, (b) caulonodin V, (c) caulonodin VI, (d) caulonodin VII; fragments of the b and c-series highlighted in red, fragments of the y and z-series highlighted in blue; peptide bonds with high frequency of breakage due to Pro are marked in red.

Identification of the sterically demanding amino acids for maintenance of the lasso fold

Several mutagenesis approaches have shown that the tolerance for substitutions in the sequence of a lasso peptide is rather high with the exceptions of some hot spots *i.e.* the amino acids involved in the macrolactam ring formation (positions 1 and 7/8/9) and the so-called plug amino acids involved in the maintenance of the lasso topology.[27–31,46] Therefore, mutagenesis of possible plug amino acids and subsequent production and stability investigations can give strong hints for involvement of certain amino acids in the entrapment of the tail. From experience with other nine ring lasso peptides, the residues needed for maintenance of the lasso fold must be rather large (K, R, F, Y or W), which leaves up to three possible residues for each of the four peptides (ESI Table S3†). Hence, a set of mutants was created for each lasso peptide, replacing the possible plugs with alanine and other residues to investigate their influence on production and stability of the resulting variants.

Analysis of the production and stability of the six caulonodin IV variants strongly suggests that Phe17 plays the major role in stabilization of the lasso fold, since the F17A substitution was not tolerated while the F17W variant was produced on wild type level (Fig. 3 ESI Table S4†) and was stable against thermal denaturation and subsequent carboxypeptidase Y digestion (ESI Fig. S2a†).

For caulonodin V, the alanine scan of the possible plugs indicated that either Tyr16 or Trp17 is responsible for the entrapment of the tail, since both variants were only detectable in trace amounts while a control mutant T15A was produced at 3% of wild type level. The subsequent exchange of these two possible plug residues with the smaller amino acid Phe (Y16F,

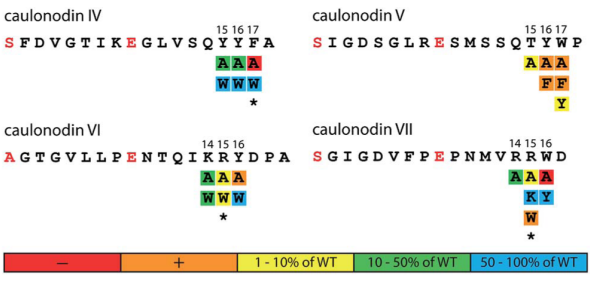


Fig. 3 Results of the mutational analysis of the plug amino acids for caulonodin IV–VII. Produced amounts of the lasso peptide variants according to UV signal integral in comparison to wild-type; color coding according to legend, asterisk marks heat stable mutants; + = detected by MS, – = not detected by MS.

W17F) also led to barely detectable amounts of lasso peptide. Only W17Y was produced at approx. 8% of wild type level (Fig. 3 ESI Table S4†). Therefore, it is very likely that both Tyr16 and Trp17 significantly contribute to the stabilization of the lasso fold and are positioned on opposite sides of the macrolactam ring.

The mutational investigation of caulonodin VI showed that while the production is lowest for the Y16A variant and Y16W is produced on wild type level (Fig. 3 ESI Table S4†), the incorporation of Trp at position 16 did not confer heat stability. Instead, the R15W variant, although only produced in small amounts, was stable against thermally induced unfolding. Subsequent carboxypeptidase Y digestion proved the resistance conferring lasso fold of this variant (ESI Fig. S2b†). On the basis of these observations it is not entirely clear if Arg15 or Tyr16 is the sterically demanding, fold-stabilizing amino acid directly below the ring. Since both residues seem to significantly participate in the fold maintenance, a structure suggestion is given where they are positioned on opposite sides of the ring (ESI Fig. S3†). This would also be in agreement with mutational studies of other lasso peptides, where the exchange of the upper plug residue with alanine could completely abolish or at least strongly diminish the peptide production.[27,28,30,31]

The plug scan for caulonodin VII revealed both variants R14A and R15A to be produced, although with a partially strongly decreased yield. The W16A variant was not produced, which suggests that Trp16 could be the most important residue for the lasso fold maintenance (Fig. 3 and Table S4†). The exchange to Tyr (W16Y) was tolerated well, but the stability was not significantly lowered. On the other hand, the incorporation of a Trp at position 15 (R15W) conferred heat stability to the peptide although the production was very low (ESI Fig. S2c†). These results suggest a very similar fold to caulonodin VI. Therefore both Arg15 and Trp16 are most likely to participate in the overall stabilization by being located on opposite sides of the macrolactam ring.

With this relatively small set of mutants it was possible to identify the possible plug amino acids of all four new caulonodins with reasonable certainty and as such to predict their fold (ESI Fig. 3†).

NMR structure elucidation of caulonodin V and fold determination of caulonodin VI

Although the mutational study provides a good idea of the three dimensional fold for all four new caulonodins, an unambiguous elucidation of the structure can only be achieved by NMR spectroscopic or X-ray crystallographic methods. Therefore, two caulonodins were selected and subjected to 1D and 2D-NMR experiments.

NMR spectra were recorded on a sample of 4.0 mg of caulonodin V dissolved in 200 μL of H_2O–D_2O (9 : 1) leading to a concentration of 10.1 mM. The samples were prepared following standard procedures (see material and methods for Experimental details). The ^{1}H spectra of caulonodin V at temperatures between 283 and 303 K with 5 K increments are shown in ESI Fig. S4.† Spectra are presented in the region 10.6–6.6 ppm for clarity and labels for the signal assignments of the amide protons are attached. Neat and well resolved single set of signals were observed and a diverse distribution of the temperature response of these signals revealed a stable lasso fold of this peptide.[47] The best signal resolution was observed at 298 K and a full signal assignment (see ESI Table S5†) of the ^{1}H signals was obtained by standard procedures.[48] A combination of DQF-COSY and NOESY produced sequential assignments (*i.e.* all αH and NH and their sequence in the backbone), and a combination of DQF-COSY and TOCSY allowed the determination of the side chains. One pure conformation was observed and full assignment of ^{1}H signals was thus obtained (NOESY spectrum: ESI Fig. S5, ^{1}H chemical shifts: ESI Table S5†). Strong NOE contacts between the NH of Ser1 and the γH of Glu9 were observed, showing an internal linkage between these two residues. Inspection of ^{1}H spectra between 283 and 303 K (ESI Fig. S4†) revealed almost no temperature dependence of the NH of Glu9 and a very weak dependence of those of Tyr16 and Trp17. Furthermore, a large number of long-range NOE contacts were observed (ESI Fig. S5†). These are the connectivity between Asp4–Tyr14, Ser5–Tyr16, Gly6–Tyr16, Leu7–Tyr16, Arg8–Tyr16, Glu9–Tyr16, Ser10–Tyr16, Ser1–Trp17, Glu9–Trp17, Leu7–Trp17, Ile2–Pro18, Ser10–Ser13 and Ser10–Gln14. All these short distances identified in the NOESY spectrum are in favor of a lasso structure of caulonodin V.

Observed NOE cross-peaks were converted into distance constraints manually and used for structure calculation as well as torsion angle constraints from unambiguous coupling constants. With these constraints, 15 minimum energy structures were obtained that are in minimal violation with the experimental data (for details see methods and ESI Table S5†).

The family of 15 structures shown in Fig. 4 represents the lasso fold of caulonodin V in aqueous solution at 25 °C (for Ramachandran plot see ESI Fig. S6†). The extra cyclic part of the peptide is threaded through the ring and thus divided into two parts, the seven membered loop and the two residue tail. This loop is up to date the second largest after microcin J25 and might therefore be suitable for epitope grafting applications.[49] In particular, the residues M11-S12-S13-Q14 form a slightly bent linear sequence that is applicable for the presentation of a peptide epitope. The two amino acids Tyr16 and Trp17 are located directly above and below the ring respectively. The structure is therefore in agreement with the postulate, which predicted Tyr16 and Trp17 to play the most important roles in stabilizing the structure, acting as the plug amino acids on opposite sides of the ring. Furthermore, the structure of caulonodin V is the first one of a lasso peptide with a 9 aa glutamate-mediated ring (29 atoms). Although the structure is stabilized in a sandwich-like manner, as it is the case for astexin-1 and microcin J25, the fold seems less stable due to the enlarged ring.

To acquire more evidence for the suggested structures of the caulonodins from the second biosynthetic machinery caulonodin VI was subjected to 1D and 2D NMR experiments. The ^{1}H spectra in temperature range 283–303 K (ESI Fig. S7†) show a chemical shift dispersion of 2.5 ppm for the amide protons, which denotes a stable secondary structure of caulonodin VI in

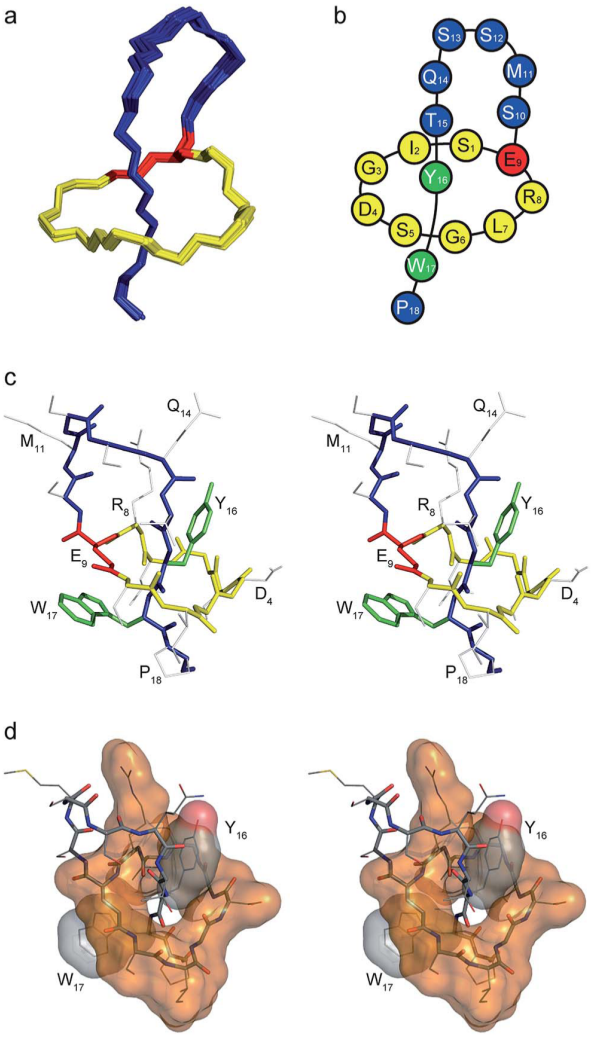


Fig. 4 3D structure of caulonodin V determined by NMR-spectroscopy. (a) Overlay of the 15 minimum energy structures, only backbone shown; (b) schematic representation of the structure; (c) relaxed eyes stereo view of the structure, backbone atoms of the aa in the ring in yellow, in the tail in blue, side chain atoms of the macrolactam ring forming aa in red, of the sterically fixating aa in green and other aa in grey; (d) interaction of steric locks in the tail and the ring, backbone and side chains shown as sticks colored by atoms, ring surface shown in orange, plug surfaces colored by atoms.

aqueous solution. The 2D spectra assured the signal assignments of the backbone and partially the side chain protons of Ala1, Gly2, Thr3, Gly4, Val5, Leu7, Glu9, Thr10, Gln12, Ile13, Lys14, Arg15, Tyr16, Asp17, and Ala19 (ESI Table S7†). Thus, long-range NOEs between NH of Ala1 and γCH_2 of Glu9 were observed verifying the isopeptide bond between these two amino acids. Further long-range NOEs (ESI Fig. S8†) were detected between the backbone NH of Arg15 and the αH of Gly2, the NH of Tyr16 and the βCH_2 of Leu7, the 2,6H and the 3,5H of Tyr16 and the αH of Gly2, the 2,6H and the 3,5H of Tyr16 and the γCH_2 of Leu7, the 2,6H and the 3,5H of Tyr16 and the βCH_2 of Glu9, and the 2,6H and the 3,5H of Tyr16 and the γCH_2 of Glu9. These contacts reveal Arg15 and Tyr16 to be the most likely residues to serve as plugs for caulonodin VI, which was already suggested by the results of the mutational analysis. However, a full structure determination was not possible, due to the strong overlay of the side chain signals. In particular, due to the reduced thermal stability of caulonodin VI, it was not possible to unambiguously assign any signals to Asn11 and the overlay in the region 0.8 to 2.5 ppm prevented full assignment of the side chain hydrogens including βH for 9 of the 19 aa. For a full 3D structure determination concerning a lasso peptide an average of about 8 to 10 unambiguous constraints per residue are needed, which is strongly dependent on clear side chain assignments. A structural calculation on the basis of the observed data would produce a 3D structure of insufficient quality and high uncertainty and is therefore not advised.

Elucidation of substrate specificity at the N-terminus

For the four new caulonodins the amino acid at position 1 is especially interesting since these are the first examples of class II lasso peptides featuring a different residue than glycine at the N-terminus. Hence, a set of mutants was created substituting the initial Ser or Ala of the best produced peptide of each cluster (caulonodin V and VI) with other amino acids of comparable size (Gly, Ala/Ser, Cys, Val and Thr) and a large one (Phe). In general agreement with previous mutagenesis studies,[27,28,30,31] the tolerance for substitution at this position was very limited (Fig. 5 ESI Table S4†). Only the Ser to Ala substitution and *vice versa* was tolerated very well. The glycine substitution lead to a strong decrease in production, indicating that the specificity of the biosynthetic machinery for the first amino acid has shifted from glycine towards the slightly larger amino acids alanine and serine. Despite the chemical and sterical similarity (Cys, Val, Thr), the other variants were barely detected. As expected, the same severe impact on production was observed for the Phe substitution, which was designed as a negative control. Although the substrate specificity of the maturation machinery for position 1 in these two clusters is still very high, it represents so far the first example of a biosynthetic machinery producing lasso peptides with two amino acids (Ala and Ser) at the initial position in similar amounts and tolerating a third amino acid (Gly) with moderately to strongly reduced production.

Identification of a lasso peptide leader motif and mutational analysis of the caulonodin leader peptides

Since *Caulobacter* sp. K31 features only two putative secondary metabolite clusters identified by antiSMASH,[50,51] it is remarkable that it contains three lasso peptide biosynthetic gene clusters with a total of seven products. These three clusters therefore present an interesting starting point for comparative bioinformatic studies. It is apparent, that the two clusters investigated in this study are very similar to each other (protein identities of over 40%), while the previously studied three precursor cluster differs significantly (protein identities of below 31%). The precursor peptides show an overall low identity

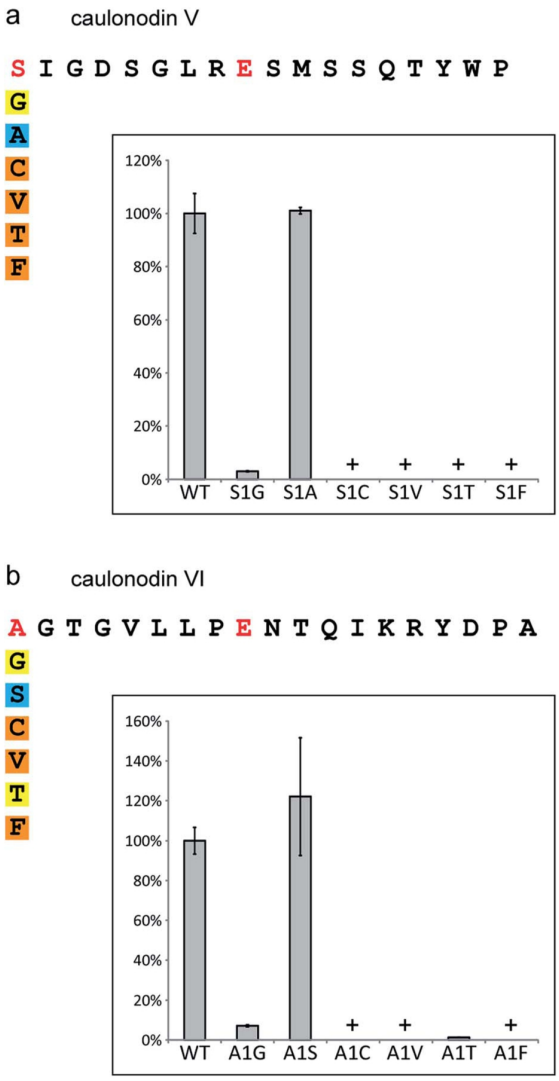


Fig. 5 Results of the mutational analysis of the specificity for position 1 in (a) the caulonodin V system and (b) the caulonodin VI system. Produced amounts of the lasso peptide variants according to UV signal integral in comparison to wild-type; color code as in Fig. 3; insets show column diagram with standard deviation error bars; + = detected by MS.

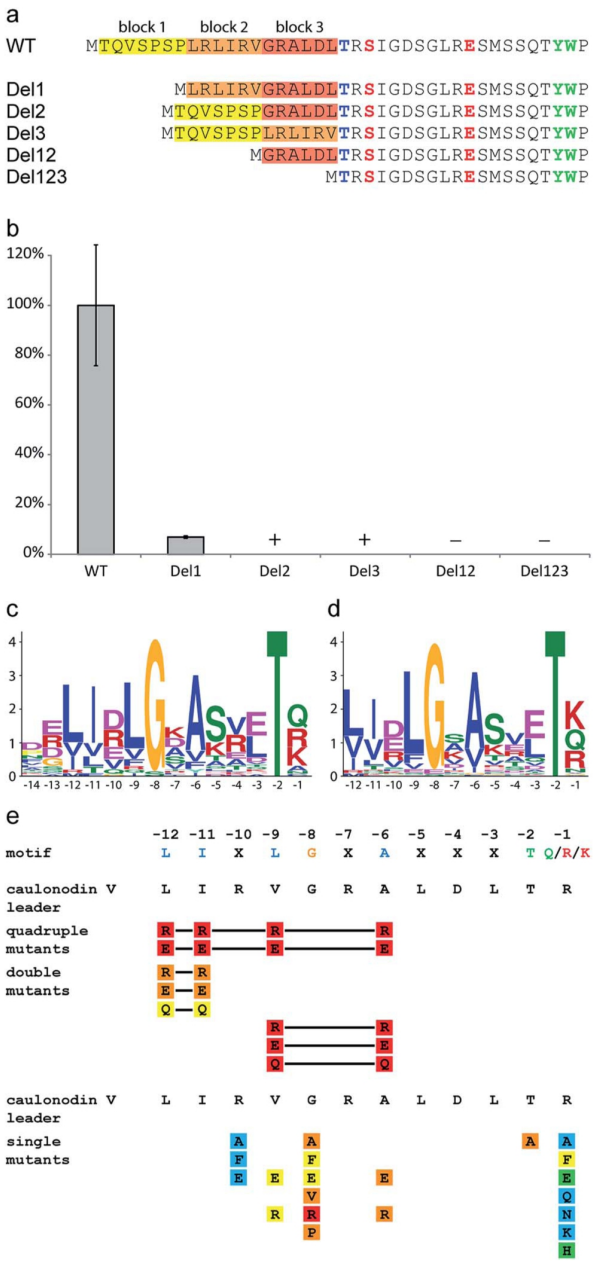


Fig. 6 Effects of the deletion, single and multiple substitution mutations of the lasso peptide leader sequence on caulonodin V production. (a) Scheme showing constructed deletions of caulonodin V leader peptide and (b) their influence on peptide production compared to WT, according to UV signal; + detected by MS, – = not detected by MS; (c) motif identified in 27 leader peptides by the MEME algorithm using all sequences from the 30 functional clusters and (d) all non-identical putative precursors from recently identified lasso gene clusters[26] as query; (e) influence of substitutions in the precursor on production of caulonodin V; color code as in Fig. 3.

of only 9% and a similarity of 16%. Interestingly these residues are mainly located in the leader peptide in a range of ten amino acids in front of the cleavage site (ESI Fig. S9†). To verify the importance of this region five deletion mutants of the precursor genes coding for caulonodin V were created omitting six or seven amino acid long parts of the leader sequence of the precursor peptide. The initial methionine and the last two amino acids of the leader peptide containing the previously investigated highly conserved Thr-2 were left untouched (Fig. 6a).[27–29,45] The results showed that only deletion of the first seven amino acids led to an amount of produced lasso peptide detectable by UV (Fig. 6b ESI Table S4†). All other deletions almost completely abolished the production. It therefore seems that the C-terminal 14 residues of the leader sequence play a major role in the interaction with the biosynthetic machinery and the processing of the precursor into the mature lasso

peptide. A potential general recognition site for the processing of lasso peptides might be located in this region.[52] It is known from other RiPPs like cyanobactins and lanthipeptides, that recognition sequences often directly precede or follow the core peptide sequence.[53,54]

Recently, several new lasso peptides have been isolated by genome mining approaches and 30 lasso peptides and their respective gene clusters are known. Although they are mainly produced by proteobacteria, three examples are from an actinomycetal origin[6,55] and therefore present the basis for a broader view. A bioinformatic analysis of the precursor peptides from all known functional lasso clusters with the MEME algorithm[56] unveiled a motif located in the C-terminal part of the leader peptide ranging from position −12 to −1 (Fig. 6c) with the consensus sequence LIxLGxAxxxTx. The same motif was identified, when all putative precursors identified from a recent genome mining study[26] were used as the query (Fig. 6d). Interestingly, this motif was not observed in the precursors of the prototypical lasso peptide microcin J25, capistruin and burhizin. On the other hand it was not only found in 24 proteobacteria but also in the three actinomyces species and it may therefore present a more or less general recognition motif in the leader sequence of lasso peptide precursors (alignment shown in ESI Fig. S10†).

In this motif, several residues are highly conserved, namely Gly-8 and Thr-2 as well as hydrophobic residues at positions −12, −11, −9 and −6. Furthermore there is some conservation of the residue −1 (Gln/Lys/Arg), which predominantly contains an amino function and the residue −10 (Asp/Arg/Glu), which is in most cases a charged residue, even though the polarity seems not to be conserved. The residues −7, −5, −4 and −3 are quite variable, although residue −5 is in general rather hydrophilic, while the residues −4 and −3 alternate between hydrophilic and hydrophobic.

To investigate the importance of certain residues in this newly discovered motif, 29 single or multiple substitution mutants were created for all highly and moderately conserved residues in the precursor gene coding for caulonodin V (Fig. 6e ESI Table S4†). The importance of the four hydrophobic residues at positions −12, −11, −9 and −6 was assessed with quadruple, double and certain single mutants, revealing a general importance of this hydrophobic patch for the *in vivo* maturation. Furthermore, these substitutions highlighted that Ala-6 is the most important residue of the four, since its single residue substitutions were barely detected. The previously reported importance of Thr-2 (ref. 45) was confirmed by the significantly decreased production of a T-2A mutant. The newly discovered conserved Gly-8 showed an almost comparable significance since all substitutions were processed in very small amounts (Ala, Phe, Glu, Val, Pro) or not at all (Arg). The two other residues probed in this approach R-10 and R-1 showed only minor influence on effective maturation. All three R-10 variants were produced on wild type level, rendering this position completely variable. The processing of the R-1 variants shows a preference of the maturation machinery towards medium or small residues (Ala, Glu, Gln, Asn, Lys, His) over large amino acids (Phe). Since this is the ultimate position of the leader peptide, it is very likely that besides Thr-2, this residue is also crucial for the recognition by the protease function containing protein B. Together with the previous studies the protease might recognize two medium sized residues best, while large residues disturb the interaction.

To further emphasize the general relevance of the identified motif, a mutational approach was designed to show an enhancing effect on the production of a lasso peptide upon restoration of the correct motif. For this, the precursor of the lasso peptide rhodanodin from *Rhodanobacter thiooxidans* LCS2 was chosen,[26] in which the motif was also identified (ESI Fig. S10†), but the central Gly-8 was substituted by Ser. This was especially intriguing as lasso peptide precursors found in other *Rhodanobacter* species conformed to the conserved motif. The exchange mutation S-8G led to a more than 10–fold production increase (ESI Table S4†), highlighting the significance of this residue within the motif.

It was shown that several conserved residues, which are clustered in the identified motif, have a strong impact on the effective processing of the precursor peptide to the mature lasso peptide *in vivo*. Some of these residues, especially Gly-8 and Ala-6 can be considered as equally important for the enzyme recognition as the previously identified Thr-2.

Conclusion

In this work, we applied genome mining techniques to broaden the scope for new lasso peptides beyond previously established boundaries by looking for unusual precursor peptide sequences. Four of these sequences are present in two clusters from the strain *Caulobacter* sp. K31, which also features a standard lasso peptide biosynthesis cluster that was recently characterized. Utilizing the engineered clusters in an *E. coli* expression system, it was possible to isolate and characterize four new lasso peptides, the caulonodins IV to VII, featuring Ser or Ala at position 1, breaking the paradigm of class II lasso peptides. A mutational study focusing on the identification of the plug amino acids, made it possible to predict the folds of all four new caulonodins with reasonable certainty. These predictions were tested for two instances, caulonodin V and VI. The 3D fold and structure elucidation by NMR techniques confirmed the postulated folds for both caulonodins and unveiled a lasso structure of caulonodin V with a seven membered loop and a two residue tail. The substrate specificity and the tolerance for substitution of the atypical amino acid at position 1 was probed for both biosynthesis clusters in a mutational approach revealing a shifted specificity from Gly to Ser and Ala. Both residues were accepted equally well, while general tolerance for substitution at this key position is very limited.

The second major part of this study focused on the leader peptide and its influence on effective processing by the biosynthetic enzymes. An initial deletion study pointed towards the C-terminal part as an important region of the leader peptide. A subsequent bioinformatics analysis using the MEME algorithm uncovered a motif with several highly conserved residues. These residues were analyzed for their influence on *in vivo* processing of the precursor peptide. Substitutions of

several residues had a significant impact on production comparable to Thr-2 exchanges, revealing their importance for effective processing. Since these residues are clustered in an area of the precursor close to the cleavage site, it is likely that they form a recognition motif for the post-translational modification enzymes. Hence, this motif may directly interact with the lasso peptide processing machinery consisting of proteins B and C in a similar fashion as it was shown for other RiPPs like lanthipeptides by shifting the equilibrium of the enzymes from an inactive to an active state.[53,57] Future studies focusing on the interaction of soluble lasso peptide processing enzymes with the precursors or co-crystallization experiments may provide further insight into the mechanism of this interaction.

Acknowledgements

Financial support from the Deutsche Forschungsgemeinschaft and the LOEWE program of the state of Hesse is gratefully acknowledged.

References

1 A. G. Poth, L. Y. Chan and D. J. Craik, *Biopolymers*, 2013, **100**, 480–491.
2 Y. Ji, S. Majumder, M. Millard, R. Borra, T. Bi, A. Y. Elnagar, N. Neamati, A. Shekhtman and J. A. Camarero, *J. Am. Chem. Soc.*, 2013, **135**, 11623–11633.
3 C. D. Deane and D. A. Mitchell, *J. Ind. Microbiol. Biotechnol.*, 2014, **41**, 315–331.
4 M. Hedvat, L. Emdad, S. K. Das, K. Kim, S. Dasgupta, S. Thomas, B. Hu, S. Zhu, R. Dash, B. A. Quinn, R. A. Oyesanya, T. P. Kegelman, U. K. Sokhi, S. Sarkar, E. Erdogan, M. E. Menezes, P. Bhoopathi, X. Y. Wang, M. G. Pomper, J. Wei, B. Wu, J. L. Stebbins, P. W. Diaz, J. C. Reed, M. Pellecchia, D. Sarkar and P. B. Fisher, *Anti-Cancer Agents Med. Chem.*, 2012, **12**, 1143–1155.
5 J. E. Velasquez and W. A. van der Donk, *Curr. Opin. Chem. Biol.*, 2011, **15**, 11–21.
6 R. D. Kersten, Y. L. Yang, Y. Q. Xu, P. Cimermancic, S. J. Nam, W. Fenical, M. A. Fischbach, B. S. Moore and P. C. Dorrestein, *Nat. Chem. Biol.*, 2011, **7**, 794–802.
7 P. G. Arnison, M. J. Bibb, G. Bierbaum, A. A. Bowers, T. S. Bugni, G. Bulaj, J. A. Camarero, D. J. Campopiano, G. L. Challis, J. Clardy, P. D. Cotter, D. J. Craik, M. Dawson, E. Dittmann, S. Donadio, P. C. Dorrestein, K. D. Entian, M. A. Fischbach, J. S. Garavelli, U. Goransson, C. W. Gruber, D. H. Haft, T. K. Hemscheidt, C. Hertweck, C. Hill, A. R. Horswill, M. Jaspars, W. L. Kelly, J. P. Klinman, O. P. Kuipers, A. J. Link, W. Liu, M. A. Marahiel, D. A. Mitchell, G. N. Moll, B. S. Moore, R. Muller, S. K. Nair, I. F. Nes, G. E. Norris, B. M. Olivera, H. Onaka, M. L. Patchett, J. Piel, M. J. Reaney, S. Rebuffat, R. P. Ross, H. G. Sahl, E. W. Schmidt, M. E. Selsted, K. Severinov, B. Shen, K. Sivonen, L. Smith, T. Stein, R. D. Sussmuth, J. R. Tagg, G. L. Tang, A. W. Truman, J. C. Vederas, C. T. Walsh, J. D. Walton, S. C. Wenzel, J. M. Willey and W. A. van der Donk, *Nat. Prod. Rep.*, 2013, **30**, 108–160.
8 T. J. Oman and W. A. van der Donk, *Nat. Chem. Biol.*, 2010, **6**, 9–18.
9 M. O. Maksimov and A. J. Link, *J. Ind. Microbiol. Biotechnol.*, 2013, DOI: 10.1007/s10295-013-1357-4.
10 K. A. Wilson, M. Kalkum, J. Ottesen, J. Yuzenkova, B. T. Chait, R. Landick, T. Muir, K. Severinov and S. A. Darst, *J. Am. Chem. Soc.*, 2003, **125**, 12475–12483.
11 K. J. Rosengren, R. J. Clark, N. L. Daly, U. Goransson, A. Jones and D. J. Craik, *J. Am. Chem. Soc.*, 2003, **125**, 12464–12474.
12 M. J. Bayro, J. Mukhopadhyay, G. V. Swapna, J. Y. Huang, L. C. Ma, E. Sineva, P. E. Dawson, G. T. Montelione and R. H. Ebright, *J. Am. Chem. Soc.*, 2003, **125**, 12382–12383.
13 K. P. Yan, Y. Li, S. Zirah, C. Goulard, T. A. Knappe, M. A. Marahiel and S. Rebuffat, *ChemBioChem*, 2012, **13**, 1046–1052.
14 S. J. Pan, J. Rajniak, W. L. Cheung and A. J. Link, *ChemBioChem*, 2012, **13**, 367–370.
15 T. A. Knappe, U. Linne, S. Zirah, S. Rebuffat, X. Xie and M. A. Marahiel, *J. Am. Chem. Soc.*, 2008, **130**, 11446–11454.
16 K. Kuznedelov, E. Semenova, T. A. Knappe, D. Mukhamedyarov, A. Srivastava, S. Chatterjee, R. H. Ebright, M. A. Marahiel and K. Severinov, *J. Mol. Biol.*, 2011, **412**, 842–848.
17 K. Adelman, J. Yuzenkova, A. La Porta, N. Zenkin, J. Lee, J. T. Lis, S. Borukhov, M. D. Wang and K. Severinov, *Mol. Cell*, 2004, **14**, 753–762.
18 M. O. Maksimov and A. J. Link, *J. Am. Chem. Soc.*, 2013, **135**, 12038–12047.
19 S. Duquesne, D. Destoumieux-Garzon, S. Zirah, C. Goulard, J. Peduzzi and S. Rebuffat, *Chem. Biol.*, 2007, **14**, 793–803.
20 T. A. Knappe, U. Linne, X. Xie and M. A. Marahiel, *FEBS Lett.*, 2010, **584**, 785–789.
21 D. F. Wyss, H. W. Lahm, M. Manneberg and A. M. Labhardt, *J. Antibiot.*, 1991, **44**, 172–180.
22 W. Weber, W. Fischli, E. Hochuli, E. Kupfer and E. K. Weibel, *J. Antibiot.*, 1991, **44**, 164–171.
23 R. A. Salomon and R. N. Farias, *J. Bacteriol.*, 1992, **174**, 7428–7435.
24 D. Frechet, J. D. Guitton, F. Herman, D. Faucher, G. Helynck, B. Monegier du Sorbier, J. P. Ridoux, E. James-Surcouf and M. Vuilhorgne, *Biochemistry*, 1994, **33**, 42–50.
25 S. Rebuffat, A. Blond, D. Destoumieux-Garzon, C. Goulard and J. Peduzzi, *Curr. Protein Pept. Sci.*, 2004, **5**, 383–391.
26 J. D. Hegemann, M. Zimmermann, S. Zhu, D. Klug and M. A. Marahiel, *Biopolymers*, 2013, **100**, 527–542.
27 M. Zimmermann, J. D. Hegemann, X. Xie and M. A. Marahiel, *Chem. Biol.*, 2013, **20**, 558–569.
28 J. D. Hegemann, M. Zimmermann, X. Xie and M. A. Marahiel, *J. Am. Chem. Soc.*, 2013, **135**, 210–222.
29 J. D. Hegemann, M. Zimmermann, S. Zhu, H. Steuber, K. Harms, X. Xie and M. A. Marahiel, *Angew. Chem., Int. Ed.*, 2014, DOI: 10.1002/anie.201309267.
30 O. Pavlova, J. Mukhopadhyay, E. Sineva, R. H. Ebright and K. Severinov, *J. Biol. Chem.*, 2008, **283**, 25589–25595.

31 T. A. Knappe, U. Linne, L. Robbel and M. A. Marahiel, *Chem. Biol.*, 2009, **16**, 1290–1298.

32 K. Severinov, E. Semenova, A. Kazakov, T. Kazakov and M. S. Gelfand, *Mol. Microbiol.*, 2007, **65**, 1380–1394.

33 J. Chiu, D. Tillett, I. W. Dawes and P. E. March, *J. Microbiol. Methods*, 2008, **73**, 195–198.

34 A. Hemsley, N. Arnheim, M. D. Toney, G. Cortopassi and D. J. Galas, *Nucleic Acids Res.*, 1989, **17**, 6545–6551.

35 M. Rance, O. W. Sorensen, G. Bodenhausen, G. Wagner, R. R. Ernst and K. Wuthrich, *Biochem. Biophys. Res. Commun.*, 1983, **117**, 479–485.

36 A. Bax and D. G. Davis, *J. Magn. Reson.*, 1985, **65**, 355–360.

37 J. Jeener, B. H. Meier, P. Bachmann and R. R. Ernst, *J. Chem. Phys.*, 1979, **71**, 4546–4553.

38 D. Marion, M. Ikura, R. Tschudin and A. Bax, *J. Magn. Reson.*, 1989, **85**, 393–399.

39 T. L. Hwang and A. J. Shaka, *J. Magn. Reson., Ser. A*, 1995, **112**, 275–279.

40 T. D. Goddard and D. J. Kneller, University of California, San Fransisco, 2007.

41 T. Herrmann, P. Guntert and K. Wuthrich, *J. Mol. Biol.*, 2002, **319**, 209–227.

42 G. Wagner, *Prog. Nucl. Magn. Reson. Spectrosc.*, 1990, **22**, 101–139.

43 *Tripos*, Tripos Inc., 1699 South Hanley Rd. St. Louis, MO 631444, 7.3 edn, 2006.

44 M. O. Maksimov, I. Pelczer and A. J. Link, *Proc. Natl. Acad. Sci. U. S. A.*, 2012, **109**, 15223–15228.

45 S. J. Pan, J. Rajniak, M. O. Maksimov and A. J. Link, *Chem. Commun.*, 2012, **48**, 1880–1882.

46 S. J. Pan and A. J. Link, *J. Am. Chem. Soc.*, 2011, **133**, 5016–5023.

47 X. Xie and M. A. Marahiel, *ChemBioChem*, 2012, **13**, 621–625.

48 K. Wüthrich, *NMR of proteins and nucleic acids*, Wiley, New York, 1986.

49 T. A. Knappe, F. Manzenrieder, C. Mas-Moruno, U. Linne, F. Sasse, H. Kessler, X. Xie and M. A. Marahiel, *Angew. Chem., Int. Ed.*, 2011, **50**, 8714–8717.

50 K. Blin, M. H. Medema, D. Kazempour, M. A. Fischbach, R. Breitling, E. Takano and T. Weber, *Nucleic Acids Res.*, 2013, **41**, W204–W212.

51 M. H. Medema, K. Blin, P. Cimermancic, V. de Jager, P. Zakrzewski, M. A. Fischbach, T. Weber, E. Takano and R. Breitling, *Nucleic Acids Res.*, 2011, **39**, W339–W346.

52 W. L. Cheung, S. J. Pan and A. J. Link, *J. Am. Chem. Soc.*, 2010, **132**, 2514–2515.

53 G. C. Patton, M. Paul, L. E. Cooper, C. Chatterjee and W. A. van der Donk, *Biochemistry*, 2008, **47**, 7342–7351.

54 E. W. Schmidt, J. T. Nelson, D. A. Rasko, S. Sudek, J. A. Eisen, M. G. Haygood and J. Ravel, *Proc. Natl. Acad. Sci. U. S. A.*, 2005, **102**, 7315–7320.

55 J. Inokoshi, M. Matsuhama, M. Miyake, H. Ikeda and H. Tomoda, *Appl. Microbiol. Biotechnol.*, 2012, **95**, 451–460.

56 T. L. Bailey, M. Boden, F. A. Buske, M. Frith, C. E. Grant, L. Clementi, J. Ren, W. W. Li and W. S. Noble, *Nucleic Acids Res.*, 2009, **37**, W202–W208.

57 T. J. Oman, P. J. Knerr, N. A. Bindman, J. E. Velasquez and W. A. van der Donk, *J. Am. Chem. Soc.*, 2012, **134**, 6952–6955.

3.6 Rational Improvement of the Affinity and Selectivity of Integrin Binding of Grafted Lasso Peptides

Julian D. Hegemann, Mariarosaria De Simone, Marcel Zimmermann, Thomas A. Knappe, Xiulan Xie, Francesco Saverio Di Leva, Luciana Marinelli, Ettore Novellino, Stefan Zahler, Horst Kessler, and Mohamed A. Marahiel, Rational Improvement of the Affinity and Selectivity of Integrin Binding of Grafted Lasso Peptides,
J. Med. Chem. **2014**, *57(13)*, 5829-5834.
doi: 10.1021/jm5004478

Author contributions:

The project was designed by Julian D. Hegemann together with Thomas A. Knappe, Horst Kessler, Marcel Zimmermann and Mohamed A. Marahiel. Generation of the MccJ25 variants and their isolation was done by Julian D. Hegemann. Integrin binding studies were done and evaluated in collaboration with Mariarosaria De Simone and Horst Kessler, while cell culture studies were performed and interpreted by Stefan Zahler. The NMR spectroscopic elucidation of the three-dimensional structure of MccJ25(RGDF) was done by Xiulan Xie and the molecular dynamics simulations were carried out by Francesco Saverio Di Leva, Luciana Marinelli and Ettore Novellino. Project coordination was done predominantly by Julian D. Hegemann and Mohamed A. Marahiel. The manuscript was written by Julian D. Hegemann and revised after thorough exchange with all parties involved in this project. The general project supervision was performed by Mohamed A. Marahiel.

Background and Summary:

Due to their extraordinary stability, genetic accessibility and their promiscuous primary structures, lasso peptides are interesting scaffolds for epitope grafting approaches. Still, so far only one study was carried out to assess the applicability of lasso peptides in this field.[65] Even though this proof-of-principle study was able to verify that a peptide epitope can be grafted into a lasso scaffold and still retains its original bioactivity, the resulting compound had only a moderate binding affinity and an even worse binding selectivity to the target integrin receptor. Based on this result, we set out to optimize the grafted lasso peptide by applying findings from the research that led to the development of cilengitide, a highly potent and selective inhibitor of the αvβ3 integrin receptor. The newly generated lasso grafts were then probed in *in vitro* integrin binding studies, revealing that MccJ25(RGDF) had a significantly improved αvβ3 binding behavior, both in regard to affinity and selectivity of the interaction. This result was further substantiated by cell culture studies that confirmed that this compound indeed conferred the desired inhibitory activity. To shed light upon the structure-activity relationship of this novel compound, its structure was elucidated by NMR spectroscopy and then employed in a molecular dynamics simulation using the co-crystal structure of a cilengitide-αvβ3 complex (PDB code 1L5G)[130] as a basis. In this way, it was revealed which moieties of MccJ25(RGDF) are important for receptor recognition and of which nature the inhibitor-receptor interactions are on a molecular level.

Journal of Medicinal Chemistry

Brief Article

pubs.acs.org/jmc

Rational Improvement of the Affinity and Selectivity of Integrin Binding of Grafted Lasso Peptides

Julian D. Hegemann,[†] Mariarosaria De Simone,[‡,§,#] Marcel Zimmermann,[†] Thomas A. Knappe,[†,∞] Xiulan Xie,[†] Francesco Saverio Di Leva,[∥] Luciana Marinelli,[∥] Ettore Novellino,[∥] Stefan Zahler,[⊥] Horst Kessler,[‡,§] and Mohamed A. Marahiel*[,†]

[†]Department of Chemistry/Biochemistry, LOEWE Center for Synthetic Microbiology, Philipps-Universität Marburg, Hans-Meerwein-Strasse 4, 35032 Marburg, Germany

[‡]Department of Chemistry, Institute for Advanced Study, Center of Integrated Protein Science (CIPSM), Technische Universität München, Lichtenbergstrasse 4, 85747 Garching, Germany

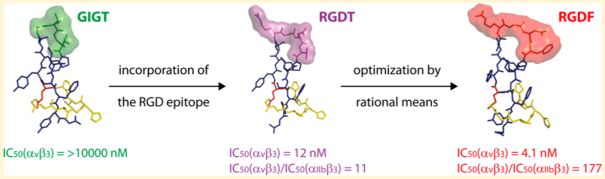
[§]Department of Chemistry, Faculty of Science, King Abdulaziz University, P.O. Box 80203, Jeddah 21589, Saudi Arabia

[∥]Department of Pharmacy, Università di Napoli "Federico II", Via D. Montesano 49, 80131 Napoli, Italy

[⊥]Department of Pharmacy, Ludwig Maximilian University, Butenandtstrasse 5-13, 81377 Munich, Germany

Supporting Information

ABSTRACT: Integrins moderate diverse important functions in the human body and are promising targets in cancer therapy. Hence, the selective inhibition of specific integrins is of great medicinal interest. Here, we report the optimization of a grafted lasso peptide, yielding MccJ25(RGDF), which is a highly potent and selective $\alpha v\beta 3$ integrin inhibitor. Furthermore, its NMR structure was elucidated and employed in a molecular dynamics approach, revealing information about the integrin binding mode and selectivity profile of MccJ25(RGDF).



INTRODUCTION

Peptides have an enormous potential for the development of compounds with specific inhibitory or antagonistic activities. Especially the easy access to a large peptide diversity through chemical and ribosomal synthesis facilitates the generation and screening of vast libraries of diverse compounds in search for desired activities.[1−4] However, despite this potential, linear peptides have the major drawback of being conformationally promiscuous and having a low long-term stability under physiological conditions.[5−8] An efficient way of combining the inherent potential of peptides with an improved stability and reduced flexibility is to generate cyclic peptides or to graft bioactive peptide epitopes into naturally occurring highly stable scaffolds.[4−11] The most prominent examples for scaffolds employed in the latter approach are probably cyclotides,[5,7,8] but also bacterial lasso peptides have been recently used for this purpose.[6]

Both cyclotides and lasso peptides share a high stability against thermal, chemical, and proteolytical degradation that is conferred by their rigid structures because of their respective folds.[5−8,12−16] While cysteine-knot containing cyclotides have the advantage that they allow the grafting of larger epitopes by synthetic means, the lasso fold can be maintained without having to rely on disulfide bridges, making them resistant to reducing conditions. This paves the way for employing specialized grafted peptide scaffolds for specific problems. Cyclotides will thus prevail when the target calls for a large peptide epitope to be incorporated, while lasso peptides could prove to be superior when small epitopes allow the functionalization of their more compact and reducing agent resistant structures. Nevertheless, the lasso fold still remains inaccessible by chemical synthesis; thus, production of these compounds requires the heterologous expression of their biosynthetic machineries in microorganisms.

These machineries consist of at least two ATP-dependent enzymes, a cysteine protease and a macrocyclase, which process the gene encoded precursor into the mature lasso peptide. This is achieved by cleaving off the precursors leader peptide and by catalyzing the formation of the threaded macrocycle between the N-terminal α-amino group and the side chain of an aspartate or glutamate residue at position 7, 8, or 9.[13−19] Still, the genetically accessible primary structures and the highly promiscuous processing enzymes allow an efficient alteration and modification of such lasso peptide scaffolds.[6,13,15,16,20−22]

Recently, we reported that the incorporation of the RGD epitope into the loop region of the lasso peptide microcin J25 (MccJ25) yielded a potent integrin antagonist, namely, MccJ25(RGD).[6] However, the intergrin binding affinity and especially the selectivity profile of MccJ25(RGD) were not wholly satisfying. In general, integrins are a superfamily of heterodimeric cell surface receptors that finely moderate a

Received: March 21, 2014
Published: June 20, 2014

Table 1. Binding Affinity of the Tested Compounds for the $\alpha v\beta 3$, $\alpha v\beta 5$, $\alpha 5\beta 1$ and $\alpha IIb\beta 3$ Integrins[a]

peptide	$\alpha v\beta 3$	$\alpha v\beta 5$	$\alpha 5\beta 1$	$\alpha IIb\beta 3$	$\alpha IIb\beta 3$/ $\alpha v\beta 3$
McJ25 (RGD)	12 ± 1.9	38 ± 4.7	830 ± 180	130 ± 15	11
McJ25 (FRGD)	21 ± 0.9	80 ± 3.5	66 ± 0.3	947 ± 150	45
McJ25 (RGDF)	4.1 ± 0.8	92 ± 10	1546 ± 0.5	725 ± 180	177
cilengitide[b]	0.2 ± 0.01	5.4 ± 1.8	10.1 ± 0.5	244 ± 10.4	1220
tirofiban[c]	-	-	-	2.4 ± 0.2	-

[a]The IC_{50} values [nM] were determined in isolated receptor binding experiments using ELISAs.[36] [b]The observed values are in good accordance with the values for cilengitide reported in the literature[6] ($IC_{50}(\alpha v\beta 3)$ = 0.9 nM, $IC_{50}(\alpha v\beta 5)$ = 25 nM, $IC_{50}(\alpha 5\beta 1)$ = 8 nM, $IC_{50}(\alpha IIb\beta 3)$ = 206 nM). [c]The observed value is in good accordance with the value for tirofiban reported in the literature[6] ($IC_{50}(\alpha IIb\beta 3)$ = 0.6 nM).

diverse array of functions in the human body; thus, the selective targeting of single integrins is important for therapeutic purposes to avoid undesired side effects. In particular the integrins $\alpha v\beta 3$ and $\alpha v\beta 5$ have emerged as promising targets for cancer treatment, as they play an important role in tumor angiogenesis and hence are crucial factors in respect to tumor growth and metastasis.[9,23−26] Previous studies on cyclic peptides led to the discovery of the cyclic pentapeptide cilengitide (cyclo[RGDf-N(Me)V-]) as a very potent and biselective integrin inhibitor for $\alpha v\beta 3$ and $\alpha v\beta 5$, which is currently in clinical trials for the treatment of lung cancer, pontine gliomas, and is, in combination with paclitaxel, under evaluation for breast cancer treatment.[9,27] In this study, we strove to apply the lessons learned during the development of cilengitide to our lasso peptide scaffold to increase its $\alpha v\beta 3$ binding affinity and selectivity. Our results highlight that a grafted lasso peptide can be easily modified in order to achieve optimal potency and selectivity toward the desired target. Furthermore, the NMR structure of the most active and selective compound, McJ25(RGDF), was elucidated and utilized in molecular dynamics (MD) simulations to rationalize its integrin binding affinity and selectivity profile. Finally, the capability of McJ25(RGDF) to inhibit integrin mediated capillary formation was proven in an in vitro assay employing human umbilical vein endothelial cells (HUVECs).

RESULTS AND DISCUSSION

During the development of cilengitide, significant improvements in the $\alpha v\beta 3$ binding behavior of cyclic peptides were achieved through N-methylation of specific amide bonds and the incorporation of a phenylalanine residue adjacent to the RGD binding motif.[9,27−31] Computational studies indeed revealed that the aromatic side chain of such an amino acid can establish multiple favorable interactions with residues comprising the receptor β subunit.[32,33] In the case of cilengitide, the best results were obtained by using D-Phe, which allows the RGD motif to adopt the optimal conformation for integrin binding.[9,30]

However, the ribosomal origin of lasso peptides does not allow the incorporation of D-amino acids or the selective N-methylation of specific peptide bonds.[9,27,30,31] Therefore, in an attempt to improve the $\alpha v\beta 3$ binding affinity and selectivity, an L-Phe was incorporated in the lasso peptide scaffold of McJ25(RGD) adjacent to the RGD binding motif.

Hence, two mutants of the pTUC202 plasmid[34,35] were generated in which the *mcjA* gene, which encodes the precursor peptide of McJ25, was altered to contain the V11F G12R I13G G14D or the G12R I13G G14D T15F quadruple mutation. The resulting McJ25 variants, McJ25(FRGD) and McJ25(RGDF), feature the integrin binding motif at the same position as the previously reported McJ25(RGD)[6] but also carry an additional phenylalanine residue on the N-terminal or C-terminal side of the peptide epitope. After initial test fermentations, the maturation of both variants was confirmed via LC−FT-MS. Large scale fermentations and isolation were performed under the previously reported conditions.[6] After purification of the McJ25 variants via HPLC, the purified lasso peptides were analyzed for their inhibitory potential against different integrin receptors.

For evaluation of the affinity and selectivity of integrin binding, enzyme-linked immunosorbent assays (ELISAs) were performed with the integrins $\alpha v\beta 3$, $\alpha v\beta 5$, $\alpha 5\beta 1$, and $\alpha IIb\beta 3$ using McJ25(FRGD) and McJ25(RGDF). As controls, these experiments were also carried out with McJ25(RGD), with cilengitide, and in the case of $\alpha IIb\beta 3$ additionally with tirofiban. The results of these experiments are summarized in Table 1.

Besides the integrins of interest, $\alpha v\beta 3$ and $\alpha v\beta 5$, the integrins $\alpha 5\beta 1$ and $\alpha IIb\beta 3$ were tested to probe the selectivity of the grafted lasso peptides. This is especially relevant for $\alpha IIb\beta 3$, as this integrin subtype is essential in the mediation of platelet aggregation, and thus, the inhibition of this receptor would cause severe side effects in vivo.[37]

The McJ25(FRGD) variant shows no noteworthy alteration in the $\alpha v\beta 3$ binding affinity compared to McJ25(RGD) but exhibits a considerable improvement of the $\alpha IIb\beta 3/\alpha v\beta 3$ binding selectivity. In contrast, McJ25(RGDF) shows not only a significant increase in the $\alpha v\beta 3$ binding affinity but also an increase of the $\alpha IIb\beta 3/\alpha v\beta 3$ binding selectivity by an order of magnitude when compared to McJ25(RGD). In this regard, it is also remarkable that McJ25(FRGD) and McJ25(RGDF) show a lower $\alpha IIb\beta 3$ binding affinity than cilengitide. Concerning the affinity toward $\alpha v\beta 5$, all grafted lasso peptides show a comparable potency while only McJ25(FRGD) shows a binding affinity for $\alpha 5\beta 1$ that is in the same range as that of cilengitide. Furthermore, the almost 2-fold decrease in $\alpha 5\beta 1$ binding affinity of McJ25(RGDF) compared to McJ25(RGD) emphasizes the high specificity of the former peptide for the $\alpha v\beta 3$ integrin. To better understand how this strong preference is achieved by the RGDF variant, its three-dimensional structure had to be elucidated.

Therefore, we set out to optimize the production of McJ25(RGDF) to improve yields and hence allow easy isolation of sufficient amounts of the compound for NMR spectroscopic studies. Such an optimization is also of great importance to allow a more economical and sustainable production of lasso peptides by fermentation. This was done in two steps. First, the fermentation medium was switched from M63 to M9 minimal medium. This medium was recently shown to allow heterologous production of lasso peptides with high yields[13−16] and led in our case also to a significant increase in lasso peptide production. Second, the effects of prolonging the duration of the fermentation from 1 day to 3 days at 37 °C

was assessed, leading again to an improvement of the overall yields. The positive effects of these alterations of the fermentation conditions are summarized in Figure 1. Applying the optimized production conditions, we were able to increase the production up to 3-fold and could isolate pure MccJ25-(RGDF) with yields of 13.5 mg/L culture.

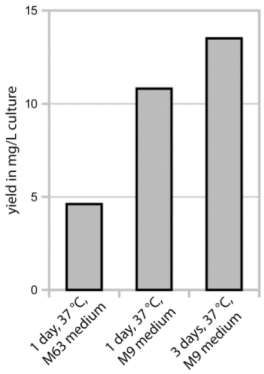


Figure 1. Optimization of MccJ25(RGDF) production in *E. coli*.

The three-dimensional structure of MccJ25(RGDF) was elucidated by employing NMR spectroscopic methods. For this, an amount of 7 mg of MccJ25(RGDF) was dissolved in methanol-d_3 and analyzed. The resulting structure confirms that MccJ25(RGDF) adopts the desired lasso fold, which is shown in Figure 2.

The superimposition of the 15 lowest energy structures (Figure 2a) demonstrates a very rigid fold of the peptides backbone and reveals that the side chains are the only parts of the compound showing some conformational flexibility. Overlaying the structures of MccJ25(RGD) and MccJ25-(RGDF) (Figure 2c) furthermore reveals that both variants adopt an almost identical fold with exception of the respective loop regions carrying the grafted peptide epitopes.

Interestingly, the peptide backbones bend in completely different directions in this area, even though the only difference in their primary structures is the amino acid residue at position 15. A comparison of the topologies of the residues 12−15 of wild type MccJ25 (Figure 2d) to MccJ25(RGD) (Figure 2e) and MccJ25(RGDF) (Figure 2f) emphasizes how the epitope grafting approach completely rearranges the upper loop region while leaving the remaining scaffold almost unaltered. A direct comparison of the surface of the peptide epitopes in MccJ25(RGD) and MccJ25(RGDF) reveals that the T15F exchange causes the RGD binding motif to adopt a flattened and less bent topology.

To investigate the binding mode of the most active grafted lasso peptide MccJ25(RGDF) into the $\alpha v\beta 3$ integrin and its $\alpha v\beta 3/\alpha 5\beta 1$ selectivity profile, molecular modeling studies were performed. However, in case of high molecular weight peptides, automated docking methods are not the best option, as they are unable to sample the large conformational space of highly flexible ligands upon binding. For this reason and with the aim to take into account the receptor flexibility and the solvent role, we first performed a manual docking of the MccJ25(RGDF) averaged NMR structure in the $\alpha v\beta 3$ crystal structure (PDB code 1L5G)[38] and then conducted an over 100 ns MD simulation. As expected, at the beginning of the MD simulation, MccJ25(RGDF) rearranges within the $\alpha v\beta 3$ RGD binding site

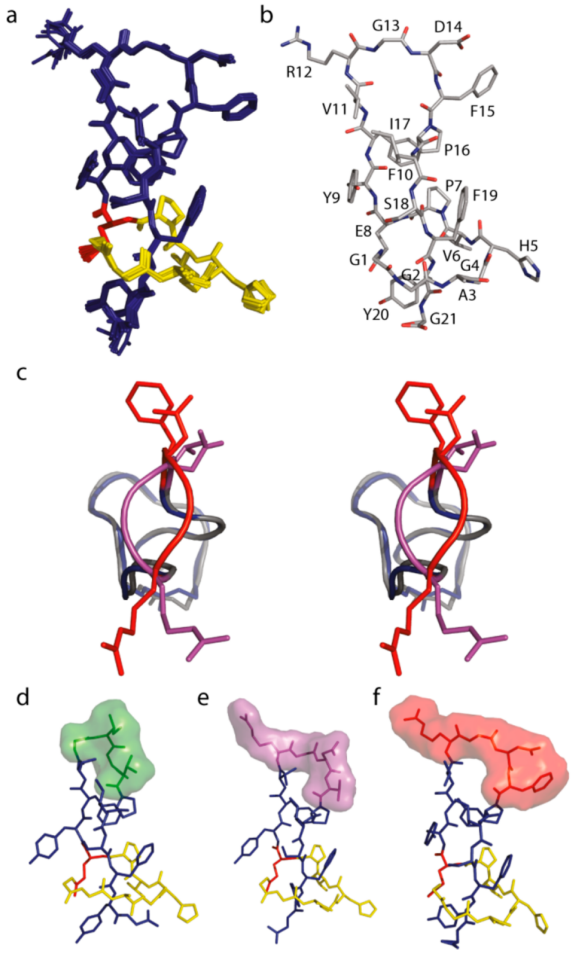


Figure 2. (a) Superimposition of the 15 lowest energy structures of MccJ25(RGDF). The tail amino acids are colored in blue, the ring forming Glu8 in red, and the remaining ring amino acids in yellow. (b) Representative average structure of MccJ25(RGDF) colored by elements (nitrogen in blue, oxygen in red, and carbon in gray). (c) Stereoview of an overlay of the structures of MccJ25(RGD) (colored in gray and magenta) and MccJ25(RGDF) (colored in blue and red) shown from above. Structures of different MccJ25 variants are shown below. The surface of the four amino acids at positions 12−15 are shown in green for (d) wild type MccJ25 (PDB code 1Q71), in magenta for (e) MccJ25(RGD) (PDB code 2MMW), and in red for (f) MccJ25(RGDF) (PDB code 2MMT).

with respect to the outcome of the manual docking, rapidly assuming a well-defined binding mode, which is conserved for the rest of the simulation (Supporting Information Figure S1). In this pose (Figure 3), the Asp14 carboxylate group coordinates the metal ion at the metal-ion-dependent adhesion site (MIDAS) and forms a hydrogen bond with the (β3)-Asn215 backbone NH amide, while the Asp14 backbone CO forms a hydrogen bond to the (β3)-Ser123 side chain. The Arg12 side chain assumes a salt bridge with (αv)-Asp218 and also establishes a cation−π interaction with (αv)-Tyr178 and a water mediated hydrogen bond with the (αv)-Thr212 hydroxyl group. Besides the RGD motif, flanking residues are also involved in the binding of MccJ25(RGDF) to $\alpha v\beta 3$. In particular, Phe15 forms a T-shape arrangement with (β3)-Tyr122 and establishes a cation−π interaction with (β3)-

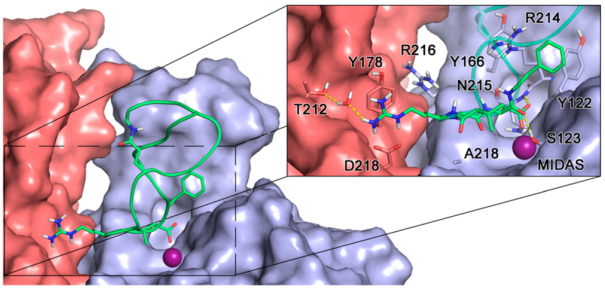


Figure 3. Binding mode of MccJ25(RGDF) at the $\alpha v\beta 3$ integrin. The lasso peptide is shown as green cartoon and sticks. The αv and $\beta 3$ receptor subunits are represented as light red and blue surface, respectively. The metal ion at the MIDAS is depicted as a purple sphere. In the inset, amino acids important for the peptide binding are highlighted as sticks. The crystal structure of the $\alpha v\beta 3$ integrin employed in the MD simulation was obtained from the protein data bank (PDB code 1L5G).[38]

Arg214. All the interactions described above are responsible for the low nanomolar $\alpha v\beta 3$ binding affinity of MccJ25(RGDF), as well as its higher potency if compared to MccJ25(RGD), in which Phe15 is replaced by a threonine residue.

In direct comparison, the binding mode of MccJ25(RGDF) closely resembles the one of cilengitide (Supporting Information Figure S2). However, despite their analogous binding modes, cilengitide is one order of magnitude more potent than MccJ25(RGDF). This is probably due to entropic effects, since small cyclic peptide ligands like cilengetide typically show reduced conformational flexibility.[9]

In order to rationalize the $\alpha v\beta 3/\alpha 5\beta 1$ selectivity profile of MccJ25(RGDF), a superposition of the $\alpha v\beta 3$ X-ray structure[38] and the $\alpha 5\beta 1$ homology model[39,40] was then accomplished (Supporting Information Figure S3). This superposition suggests that the substitution of (αv)-Thr212 with the bulkier (α5)-Gln221 generates some steric hindrance to the binding of the long Arg12 side chain. Moreover, the substitution of the (β3)-Arg214 residue with (β1)-Gly217 could further affect the $\alpha 5\beta 1$ binding affinity of MccJ25(RGDF), since the latter residue cannot interact with the Phe15 side chain. In summary, the MD simulation yielded valuable information that helps to explain the experimental data obtained in the ELISAs on a molecular level.

Still, even though the ELISAs clearly show the improved $\alpha v\beta 3$ binding affinity and selectivity of MccJ25(RGDF) in comparison to MccJ25(RGD), there is also need for the proof that the novel grafted lasso peptide has the desired inhibitory activity in qualitative cell assays. To assess the inhibitory activity on the tube formation of HUVECs, these cells were treated with different concentrations of MccJ25(RGDF) and their tube formation behavior was monitored. As control, these experiments were repeated with MccJ25(RGD), with cilengitide, and without the addition of any of these compounds. Furthermore, cell adhesion assays were performed on HUVECs with the aforementioned compounds in a likewise manner. The results of these studies are shown in Figure 4.

As can be seen, all three compounds are capable of inhibiting the tube formation and the cell adhesion in a dose-dependent manner in the tested concentration range of 3−30 μM, with cilengtide being only slightly more efficient than the grafted lasso peptides.

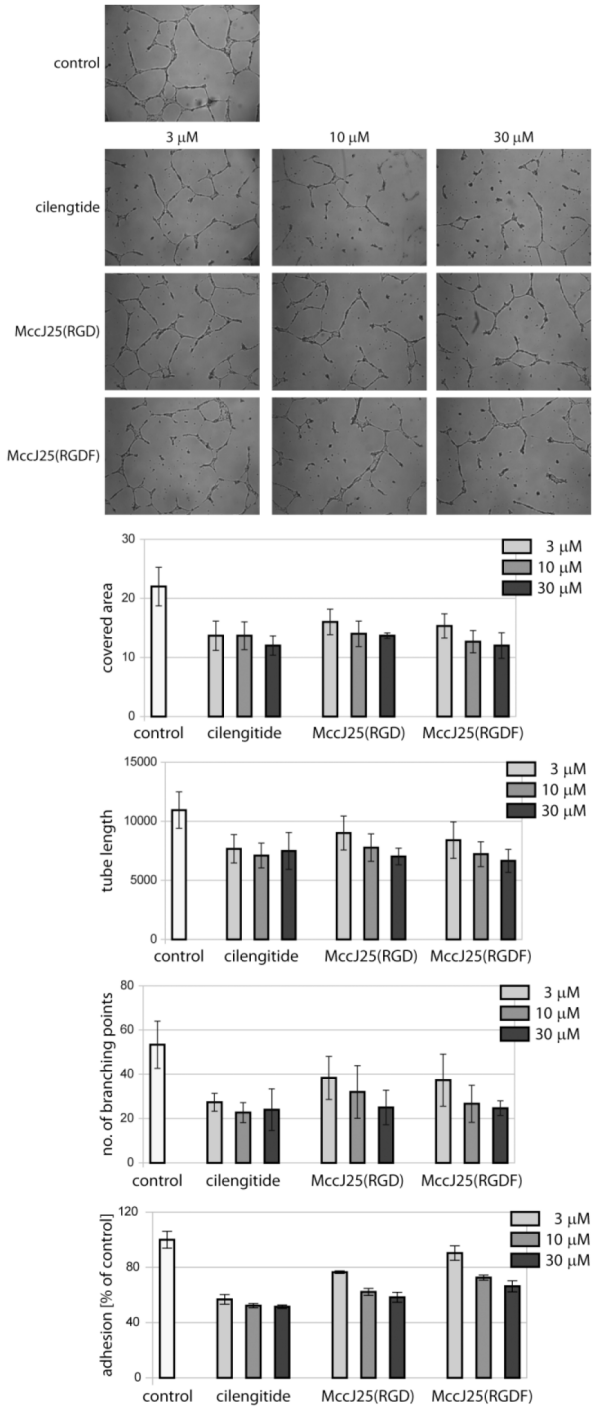


Figure 4. Inhibitory effects of the different compounds on the capillary formation and cell adhesion of HUVECs. Microscopic images of tube formation of HUVECs on Matrigel substrate is shown on top for a control experiment and in the middle for the experiments with cilengitide, MccJ25(RGD), and MccJ25(RGDF) at 3, 10, and 30 μM. The graphs depict the results of the analysis of the area covered by tubes, the tube length, and the network complexity (expressed by the number of observed branching points) under the different experimental conditions. The graph at the bottom shows the results of the cell adhesion assays.

CONCLUSION

In this study, we successfully show the improvement of a grafted lasso peptide for binding to a target integrin receptor by rational means and show its ability to inhibit the tube formation and cell adhesion of HUVECs. In this way, our new McJ25 variant, McJ25(RGDF), exhibits an increased $\alpha v\beta 3$ binding affinity and selectivity compared to the previously reported McJ25(RGD).[6] This was achieved by employing lessons learned from the research that ultimately led to the development of cilengitide[9,27−31] and thus proves that information gathered from studies with linear and circular synthetic peptides can be transferred to lasso peptide scaffolds as well. Additionally, the lasso peptide production was increased 3-fold, allowing the isolation of our improved integrin antagonist with a yield of 13.5 mg/L culture. Finally, the three-dimensional structure of McJ25(RGDF) was elucidated by NMR and used in an MD simulation for probing its $\alpha v\beta 3$ binding mode. This led to a better understanding of how the incorporation of an additional T15F mutation improved integrin binding affinity and selectivity.

In conclusion, our previous study[6] was a proof of principle that highlighted the possibility of using lasso peptides for epitope grafting efforts and yielded the integrin inhibitor McJ25(RGD), while this new study demonstrates that it is also possible to employ rational design to further improve and optimize such a scaffold toward the generation of a more potent and more selective bioactive compound. Nonetheless, the results presented here could well be the limit of optimization that can be achieved by simple genetic manipulations, as crucial factors that yielded the highly active and selective cilengitide were the incorporation of an unnatural amino acid and the chemical modification of a peptide bond. Therefore, future studies aimed at improving a lasso peptide derived integrin antagonist should also consider the use of systems that allow the incorporation of unnatural amino acids into peptides in vivo and the chemical modification of the isolated lasso peptide itself. Such additional optimization efforts were not assessed so far and could provide supplementary tools for the grafting of bioactive peptide epitopes into lasso peptide scaffolds.

ASSOCIATED CONTENT

Supporting Information

Experimental procedures, media composition, oligonucleotide primer sequences, HUVEC assay data, and details of the NMR spectroscopic structure elucidation and the MD simulation. This material is available free of charge via the Internet at http://pubs.acs.org.

Accession Codes

The NMR structure of McJ25(RGDF) was deposited at the Protein Data Bank (PDB) and is accessible via the PDB code 2MMT. The crystal structure of the $\alpha v\beta 3$ integrin in complex with cilengetide that was used for the docking experiment is accessible via the PDB code 1L5G. The NMR structure of McJ25(RGD) is accessible via the PDB code 2MMW, while the NMR structure of wild type McJ25 has the PDB code 1Q71.

AUTHOR INFORMATION

Corresponding Author

*Phone: +49 (0) 6421/2825722. E-mail: marahiel@staff.uni-marburg.de.

Present Addresses

[#]M.D.S.: Istituto Di Fisiologia Clinica (IFC-CNR), Via Moruzzi 1, Pisa 56124, Italy.

[∞]T.A.K.: Ascendis Pharma GmbH, Im Neuenheimer Feld 584, 69120 Heidelberg, Germany.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

Financial support from the Deutsche Forschungsgemeinschaft and the LOEWE program of the State of Hesse, Germany, is gratefully acknowledged. H.K. is also thankful for financial support by the Center of Integrated Protein Science Munich (CIPSM) and by the King Abdulaziz University KAU (Grant HiCi/25-3-1432).

ABBREVIATIONS USED

McJ25, microcin J25; HUVEC, human umbilical vein endothelial cell; LC−FT-MS, liquid chromatography−Fourier transform mass spectrometry; ELISA, enzyme-linked immunosorbent assay; NMR, nuclear magnetic resonance; MD, molecular dynamics; MIDAS, metal ion dependent adhesion site

REFERENCES

(1) Sato, A. K.; Viswanathan, M.; Kent, R. B.; Wood, C. R. Therapeutic peptides: technological advances driving peptides into development. *Curr. Opin. Biotechnol.* **2006**, *17*, 638−642.

(2) Antosova, Z.; Mackova, M.; Kral, V.; Macek, T. Therapeutic application of peptides and proteins: parenteral forever? *Trends. Biotechnol.* **2009**, *27*, 628−635.

(3) Pazgier, M.; Liu, M.; Zou, G.; Yuan, W.; Li, C.; Li, C.; Li, J.; Monbo, J.; Zella, D.; Tarasov, S. G.; Lu, W. Structural basis for high-affinity peptide inhibition of p53 interactions with MDM2 and MDMX. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 4665−4670.

(4) Young, T. S.; Young, D. D.; Ahmad, I.; Louis, J. M.; Benkovic, S. J.; Schultz, P. G. Evolution of cyclic peptide protease inhibitors. *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108*, 11052−11056.

(5) Ji, Y.; Majumder, S.; Millard, M.; Borra, R.; Bi, T.; Elnagar, A. Y.; Neamati, N.; Shekhtman, A.; Camarero, J. A. In vivo activation of the p53 tumor suppressor pathway by an engineered cyclotide. *J. Am. Chem. Soc.* **2013**, *135*, 11623−11633.

(6) Knappe, T. A.; Manzenrieder, F.; Mas-Moruno, C.; Linne, U.; Sasse, F.; Kessler, H.; Xie, X.; Marahiel, M. A. Introducing lasso peptides as molecular scaffolds for drug design: engineering of an integrin antagonist. *Angew. Chem., Int. Ed.* **2011**, *50*, 8714−8717.

(7) Poth, A. G.; Chan, L. Y.; Craik, D. J. Cyclotides as grafting frameworks for protein engineering and drug design applications. *Biopolymers* **2013**, *100*, 480−491.

(8) Chan, L. Y.; Gunasekera, S.; Henriques, S. T.; Worth, N. F.; Le, S. J.; Clark, R. J.; Campbell, J. H.; Craik, D. J.; Daly, N. L. Engineering pro-angiogenic peptides using stable, disulfide-rich cyclic scaffolds. *Blood* **2011**, *118*, 6709−6717.

(9) Mas-Moruno, C.; Rechenmacher, F.; Kessler, H. Cilengitide: the first anti-angiogenic small molecule drug candidate design, synthesis and clinical evaluation. *Anti-Cancer Agents Med. Chem.* **2010**, *10*, 753−768.

(10) Biron, E.; Chatterjee, J.; Ovadia, O.; Langenegger, D.; Brueggen, J.; Hoyer, D.; Schmid, H. A.; Jelinek, R.; Gilon, C.; Hoffman, A.; Kessler, H. Improving oral bioavailability of peptides by multiple N-methylation: somatostatin analogues. *Angew. Chem., Int. Ed.* **2008**, *47*, 2595−2599.

(11) Conibear, A. C.; Bochen, A.; Rosengren, K. J.; Stupar, P.; Wang, C.; Kessler, H.; Craik, D. J. The cyclic cystine ladder of theta-defensins as a stable, bifunctional scaffold: a proof-of-concept study using the integrin-binding RGD motif. *ChemBioChem* **2014**, *15*, 451−459.

(12) Salomon, R. A.; Farias, R. N. Microcin 25, a novel antimicrobial peptide produced by *Escherichia coli*. *J. Bacteriol.* **1992**, *174*, 7428−7435.
(13) Hegemann, J. D.; Zimmermann, M.; Xie, X.; Marahiel, M. A. Caulosegnins I−III: a highly diverse group of lasso peptides derived from a single biosynthetic gene cluster. *J. Am. Chem. Soc.* **2013**, *135*, 210−222.
(14) Hegemann, J. D.; Zimmermann, M.; Zhu, S.; Klug, D.; Marahiel, M. A. Lasso peptides from proteobacteria: genome mining employing heterologous expression and mass spectrometry. *Biopolymers* **2013**, *100*, 527−542.
(15) Zimmermann, M.; Hegemann, J. D.; Xie, X.; Marahiel, M. A. The astexin-1 lasso peptides: biosynthesis, stability, and structural studies. *Chem. Biol.* **2013**, *20*, 558−69.
(16) Hegemann, J. D.; Zimmermann, M.; Zhu, S.; Steuber, H.; Harms, K.; Xie, X.; Marahiel, M. A. Xanthomonins I−III: a new class of lasso peptides with a seven-residue macrolactam ring. *Angew. Chem., Int. Ed.* **2014**, *53*, 2230−2234.
(17) Yan, K. P.; Li, Y.; Zirah, S.; Goulard, C.; Knappe, T. A.; Marahiel, M. A.; Rebuffat, S. Dissecting the maturation steps of the lasso peptide microcin J25 in vitro. *ChemBioChem* **2012**, *13*, 1046−52.
(18) Maksimov, M. O.; Pelczer, I.; Link, A. J. Precursor-centric genome-mining approach for lasso peptide discovery. *Proc. Natl. Acad. Sci. U.S.A.* **2012**, *109*, 15223−15228.
(19) Maksimov, M. O.; Link, A. J. Discovery and characterization of an isopeptidase that linearizes lasso peptides. *J. Am. Chem. Soc.* **2013**, *135*, 12038−12047.
(20) Pavlova, O.; Mukhopadhyay, J.; Sineva, E.; Ebright, R. H.; Severinov, K. Systematic structure−activity analysis of microcin J25. *J. Biol. Chem.* **2008**, *283*, 25589−25595.
(21) Knappe, T. A.; Linne, U.; Robbel, L.; Marahiel, M. A. Insights into the biosynthesis and stability of the lasso peptide capistruin. *Chem. Biol.* **2009**, *16*, 1290−1298.
(22) Pan, S. J.; Link, A. J. Sequence diversity in the lasso peptide framework: discovery of functional microcin J25 variants with multiple amino acid substitutions. *J. Am. Chem. Soc.* **2011**, *133*, 5016−5023.
(23) Avraamides, C. J.; Garmy-Susini, B.; Varner, J. A. Integrins in angiogenesis and lymphangiogenesis. *Nat. Rev. Cancer* **2008**, *8*, 604−617.
(24) Desgrosellier, J. S.; Cheresh, D. A. Integrins in cancer: biological implications and therapeutic opportunities. *Nat. Rev. Cancer* **2010**, *10*, 9−22.
(25) Goodman, S. L.; Picard, M. Integrins as therapeutic targets. *Trends Pharmacol. Sci.* **2012**, *33*, 405−412.
(26) Kapp, T. G.; Rechenmacher, F.; Sobahi, T. R.; Kessler, H. Integrin modulators: a patent review. *Expert Opin. Ther. Pat.* **2013**, *23*, 1273−1295.
(27) Dechantsreiter, M. A.; Planker, E.; Mathä, B.; Lohof, E.; Hölzemann, G.; Jonczyk, A.; Goodman, S. L.; Kessler, H. N-Methylated cyclic RGD peptides as highly active and selective alpha(V)beta(3) integrin antagonists. *J. Med. Chem.* **1999**, *42*, 3033−3040.
(28) Aumailley, M.; Gurrath, M.; Müller, G.; Calvete, J.; Timpl, R.; Kessler, H. Arg-Gly-Asp constrained within cyclic pentapeptides. Strong and selective inhibitors of cell adhesion to vitronectin and laminin fragment P1. *FEBS Lett.* **1991**, *291*, 50−54.
(29) Gurrath, M.; Müller, G.; Kessler, H.; Aumailley, M.; Timpl, R. Conformation/activity studies of rationally designed potent anti-adhesive RGD peptides. *Eur. J. Biochem.* **1992**, *210*, 911−921.
(30) Haubner, R.; Gratias, R.; Diefenbach, B.; Goodman, S. L.; Jonczyk, A.; Kessler, H. Structural and functional aspects of RGD-containing cyclic pentapeptides as highly potent and selective integrin $\alpha V\beta 3$ antagonists. *J. Am. Chem. Soc.* **1996**, *118*, 7461−7472.
(31) Chatterjee, J.; Ovadia, O.; Zahn, G.; Marinelli, L.; Hoffman, A.; Gilon, C.; Kessler, H. Multiple N-methylation by a designed approach enhances receptor selectivity. *J. Med. Chem.* **2007**, *50*, 5878−5881.
(32) Marinelli, L.; Lavecchia, A.; Gottschalk, K. E.; Novellino, E.; Kessler, H. Docking studies on alphavbeta3 integrin ligands: pharmacophore refinement and implications for drug design. *J. Med. Chem.* **2003**, *46*, 4393−4404.
(33) Mas-Moruno, C.; Beck, J. G.; Doedens, L.; Frank, A. O.; Marinelli, L.; Cosconati, S.; Novellino, E.; Kessler, H. Increasing $\alpha v\beta 3$ selectivity of the anti-angiogenic drug cilengitide by N-methylation. *Angew. Chem., Int. Ed.* **2011**, *50*, 9496−9500.
(34) Solbiati, J. O.; Ciaccio, M.; Farias, R. N.; Gonzalez-Pastor, J. E.; Moreno, F.; Salomon, R. A. Sequence analysis of the four plasmid genes required to produce the circular peptide antibiotic microcin J25. *J. Bacteriol.* **1999**, *181*, 2659−2662.
(35) Solbiati, J. O.; Ciaccio, M.; Farias, R. N.; Salomon, R. A. Genetic analysis of plasmid determinants for microcin J25 production and immunity. *J. Bacteriol.* **1996**, *178*, 3661−3663.
(36) Frank, A. O.; Otto, E.; Mas-Moruni, C.; Schiller, H. B.; Marinelli, L.; Cosconati, S.; Bochen, A.; Vossmeyer, D.; Zahn, G.; Stragies, R.; Novellino, E.; Kessler, H. Conformational control of integrin-subtype selectivity in isoDGR peptide motifs: a biological switch. *Angew. Chem., Int. Ed.* **2010**, *49*, 9278−9281.
(37) Bennett, J. S. Structural biology of glycoprotein IIb-IIIa. *Trends Cardiovasc. Med.* **1996**, *6*, 31−36.
(38) Xiong, J.-P.; Stehle, T.; Zhang, R.; Joachimiak, A.; Frech, M.; Goodman, S. L.; Arnaout, M. A. Crystal structure of the extracellular segment of integrin $\alpha v\beta 3$ in complex with an Arg-Gly-Asp ligand. *Science* **2002**, *296*, 151−155.
(39) Marinelli, L.; Meyer, A.; Heckmann, D.; Lavecchia, A.; Novellino, E.; Kessler, H. Ligand binding analysis for human $\alpha 5\beta 1$ integrin: strategies for designing new $\alpha 5\beta 1$ integrin antagonists. *J. Med. Chem.* **2005**, *48*, 4204−4207.
(40) Heckmann, D.; Meyer, A.; Marinelli, L.; Zahn, G.; Strategies, R.; Kessler, H. Probing integrin selectivity: rational design of highly active and selective ligands for the alpha5beta1 and alphavbeta3 integrin receptor. *Angew. Chem., Int. Ed.* **2007**, *46*, 3571−3574.

4. Discussion

4.1 Breaking the Paradigms of Lasso Peptide Research

At the beginning, the target of this work was the genome mining driven discovery, isolation and characterization of novel lasso peptides. Back then, capistruin was still the only lasso peptide discovered in this way and the repetition of such a discovery was hindered by several inherent problems. These were predominantly the poor homologous production of lasso peptides by newly chosen strains that carried lasso peptide biosynthetic gene clusters, the poor heterologous production when cloning these clusters into a plasmid and subsequently expressing them in *E. coli* and the inexperience in regard to the criteria for choosing putative lasso peptide biosynthetic gene clusters for investigation in this way. Because of this, five years had passed since the discovery of capistruin before the next example of successful genome mining of a lasso peptide was reported in this group.[22] In the meantime, research was focused on obtaining deeper knowledge about how lasso peptides can be efficiently produced in and isolated from *E. coli*, about which sources of lasso peptide biosynthetic gene clusters are most promising for heterologous production and about what conditions can positively or negatively effect the lasso peptide production in general. Alongside our own research, further intriguing information was obtained by studies that were published in the meanwhile. Especially works of Link and co-workers[64,125] steered us in the right direction in this regard. The ideas derived from these studies ultimately led us to establish methods to significantly increase the heterologous production of lasso peptides in *E. coli* by simple genetic alterations of the corresponding biosynthetic gene clusters. By these means, we were finally able to report the isolation of three new lasso peptides, caulosegnins I-III, by a directed genome mining approach.[22] Based on this success, we revisited all the biosynthetic gene clusters that had been previously cloned into expression vectors in our group, which before showed no or only trace lasso peptide production in *E. coli*. By employing our new set of genetic tools, two things became apparent. Firstly, the genetic modification of the intergenic region in these clusters seems to be almost always beneficial for overall lasso peptide production and of general applicability. Secondly, when using *E. coli* as a heterologous host, the biosynthetic gene clusters from proteobacterial origin work best. These findings led to a comprehensive genome mining study that culminated in the description and characterization of twelve novel lasso peptides.[37] In general, the work done in this group in the recent years greatly increased the number of both known lasso peptides and known lasso peptide biosynthetic gene clusters as emphasized by Figure 4.1. Nevertheless, the results obtained in our recent publications about lasso peptides[22,25,26,37] also subtly hint at another still persisting problem. Namely, that gene clusters from organisms not belonging to the phylum Proteobacteria often still

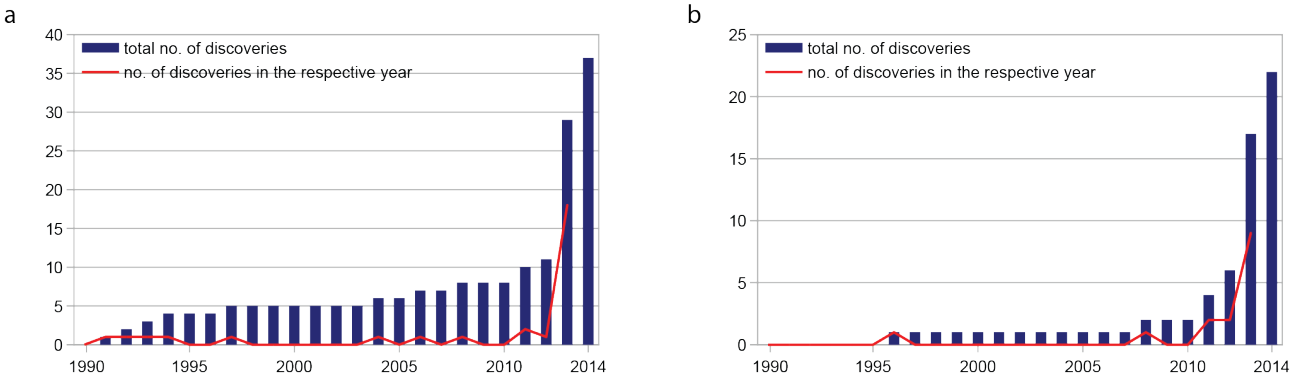


Figure 4.1. Discoveries of (a) lasso peptides and (b) functional lasso peptide biosynthetic gene clusters.[17,18,22-26,28-30,32,37,40,41,44,46,48,55,57] For the left graph, practically identical lasso peptides belonging to the siamycin or RES-701 families were counted as one single entity, even though they were rediscovered several times. As 2014 is still in progress, the red line was not expanded to this year.

cause problems when expressed heterologously in *E. coli* and that at the same time the host organisms of these gene clusters are not producing the corresponding lasso peptides homologously. Thus, even though the problem of isolating proteobacterial lasso peptides appears to be solved, a rich reservoir of novel and interesting biosynthetic systems remains currently more or less inaccessible for us. This furthermore explains why all lasso peptides described in recent publications[22,25,26,37] of our group were of proteobacterial origin.

Nonetheless, the studies presented here not only yielded a plethora of novel lasso peptides, but also vastly increased the general knowledge about these compounds and their physico-chemical properties. Especially in regard to the paradigms established before this work, new information was obtained that either put prior assumptions in question or completely overhauled these notions. Namely, these were as follows:

- All lasso peptides have an exceptional instrinsic thermal stability.
- Only the biggest amino acids are able to fulfill the role as plug amino acids.
- Only lasso peptides with eight- or nine-residue macrolactam rings can be produced.
- All lasso peptides carry either a glycine or cysteine residue at position 1 of their sequence.
- The only conserved essential residue in a lasso peptide leader sequence is the threonine at the penultimate position.
- All lasso peptides are antimicrobial agents.

These ideas, and how they were proven wrong, will be discussed in the subsequent sections.

4.1.1 The Thermal Stability of Lasso Peptides

The notion of an intrinsic thermal stability of lasso peptides goes back to the isolation of MccJ25.[32] There, it was reported that even after autoclaving, MccJ25 still exhibited its antimicrobial activity, revealing the tremendous heat resistance of this compound. Due to the presence of the general lasso fold and the basic similarities of all lasso peptides, such a behavior was assumed to be found for all members of this natural compound family. It came as quite a surprise that this assumption held not true for caulosegnins I and III. Indeed, the thermal sensitivity of caulosegnin I was only uncovered upon the first attempts of purifying this compound. This was due to the fact that after the first round of HPLC, the solvent was removed under reduced pressure in a water bath at 40 °C. Even though only the fractions corresponding to a single distinct peak were pooled and evaporated, the subsequent second HPLC purification step suddenly again revealed the presence of two well-resolved, distinct peaks. What was even more peculiar was that both peaks belonged to compounds with the exact same molecular mass. The combination of all of these observations led to the assumption that caulosegnin I most likely was partially thermally unthreaded upon the solvent evaporation at 40 °C. As a consequence, the solvent was subsequently removed more carefully and afterwards small samples of the obtained pure lasso peptide were exposed to different temperatures for different time intervals. Ensuing LC-MS analysis confirmed our initial hypothesis and revealed the appearance of a second peak featuring the same molecular mass but a different retention time (Figure 4.2).

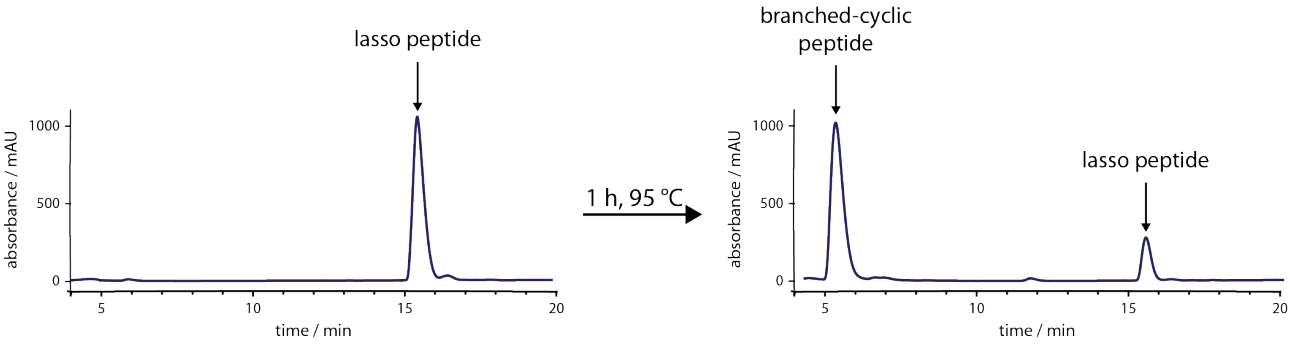


Figure 4.2. Thermally induced unthreading of caulosegnin I.[22] Even the incubation for only 1 h at 95 °C causes the almost complete unthreading of this heat sensitive lasso peptide.

Furthermore, it was shown that the intensity of this new peak directly correlated to the tested temperature and the duration of the incubation.[22] Since then, the assessment of the thermal stability became one of the benchmarks for characterizing a newly isolated lasso peptide in our group. By this, we were able to report several instances of thermally sensitive lasso peptides, namely astexin-1, caulonodin IV, caulonodin V, caulonodin VI, caulonodin VII, caulosegnin I,

caulosegnin III, rhodanodin, rubrivinodin, sphingonodin II, sphingopyxin I and syanodin I.[22,25,37,47] This is of especial interest for astexin-1, as in light of these observations, it becomes likely that the wrongfully elucidated structure of astexin-1 that was presented by Link and co-workers[30] could possibly be traced back to a thermally induced unthreading during the isolation and purification procedures of this compound and maybe even during the NMR measurements themselves. Testing the thermal stability of a new lasso peptide straight away, thus can help to avoid problems during the isolation and characterization of such a compound and can help to adjust the conditions for NMR analysis in advance to avoid unthreading during the measurements. Hence, the observation that not all lasso peptides necessarily have a high thermal stability not only overhauls a wrong belief about these compounds, but is also an important factor to consider when working with these natural products.

Additionally, our studies also helped to unveil criteria that decide upon the thermal stability or sensitivity of a lasso peptide. In this regard, the present ring size in combination with the lower plug amino acid seem to be important factors that decide how a compound reacts at elevated temperatures. While the very small seven-residue macrolactam rings are able to form a heat stable lasso fold with all lower plug amino acids bigger than serine,[26] even a tyrosine residue can in some cases be inadequate for stabilizing the fold of a nine-residue ring lasso peptide at increased temperatures.[22,25] Yet, the ring size and the character of the lower plug amino acid alone are insufficient parameters to predict the thermal stability of a lasso peptide as the comparison of caulosegnins II and III emphasizes.[22] While both lasso peptides have a nine-residue macrolactam ring, use a tyrosine residue as lower plug amino acid and have a similar amino acid composition, only caulosegnin II is heat stable. The definite reason for this difference is not known so far and is still a subject of research. Nevertheless, it is known that the distance of the lower plug amino acid to the macrolactam ring itself seems to already affect the thermal stability of a lasso peptide. This was shown for both the capistruin and caulosegnin I scaffolds, as mutational studies revealed that amino acids that are generally suitable for maintaining the lasso fold can loose this ability if positioned farther away from the point where the tail threads the macrolactam ring. In the case of caulosegnin I, isoleucine was shown to be able to maintain the fold when replacing the naturally occurring plug amino acid E16, while the mutation E16A resulted in a mere branched-cyclic peptide, even though another isoleucine residue was present at position 17 of the caulosegnin I scaffold.[22] For capistruin, it was shown that the simultaneous alanine exchange of the lower plug amino acid R15 and the adjacent F16 transformed this heat stable lasso peptide in a heat sensitive compound, where the remaining bulky F18 residue is still able to maintain the lasso fold at room temperature, but cannot stop thermally induced unthreading of the lasso fold.[36] Hence, this

capistruin variant was the first reported heat sensitive lasso peptide, while caulosegnin I was the first naturally occurring lasso peptide reported that showed a temperature sensitive behavior. In general, investigations of the thermal stability can give an insight on how stable and rigid a lasso scaffold is, ranging from a rather loose arrangement like the mentioned R15A F16A capistruin variant to the very tight lasso fold in xanthomonin II, where even amino acids in some distance to the naturally occurring plug amino acid can maintain a thermally stable fold.[26,36] Interestingly, it was also shown that the thermal properties of a lasso peptide can be inverted by rational means. As such, we have reported that the thermally sensitive astexin-1 can be transformed into a heat stable lasso peptide by introducing the single point mutation F15W.[25] This proves that both thermal stability and sensitivity are subject to change and suggests that this fact could be employed by organisms in dependence of their needs, e.g. for generating compounds that degrade at certain temperatures, allowing them to be used as temperature dependent switches. On the other hand, probing how easy a heat sensitive lasso peptide degrades, could reveal that an organism is adopted to a life in a temperature range where the thermally induced unthreading of a lasso peptide is so slow that it is of no relevance to the producing strain.

4.1.2 The Plug Amino Acids of Lasso Peptides

In addition to revealing the existence of heat sensitive lasso peptides in nature, caulosegnin I also showed that smaller amino acids as formerly believed can act as plug amino acids. In prior studies, especially in cases of the mutational analyses of the MccJ25 and capistruin scaffolds,[36,62] it seemed like only the biggest and bulkiest amino acid side chains, namely arginine, phenylalanine, tyrosine and tryptophan, were suitable for the sterical maintenance of the lasso fold. Again, caulosegnin I was the first lasso peptide proving this statement to be wrong, as it was revealed that it used a surprisingly small glutamate residue as lower plug amino acid.[22] Encouraged by this observation, we looked further into the nature of amino acids that could act as a plug. In case of caulosegnin I, we saw that the lower limit for a plug amino acid in this eight-residue macrolactam ring carrying lasso peptide was asparagine. This amino acid was only barely able to maintain the fold under the expression and isolation conditions, as it was observed that the majority of the corresponding lasso peptide variant was already unthreaded upon the actual LC-MS analysis (Figure 4.3).[26]

Reducing the size of the macrolactam ring, unsurprisingly also reduces the needed size of an amino acid side chain for stabilizing the lasso fold. Following this reasoning, it makes sense that the xanthomonin II scaffold can be maintained in a completely heat stable manner by all amino acids that exceed serine in size.[26] Of course, as demonstrated by the difference in thermal stability properties of the very similar lasso peptides caulsegnin II and III,[22] the plug amino acid alone is not

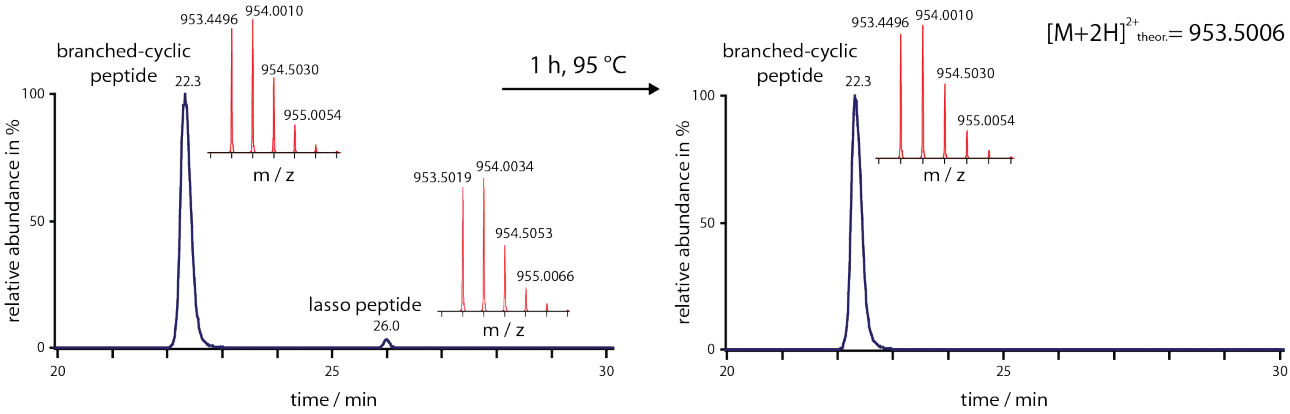


Figure 4.3. Thermal treatment of an extract of caulosegnin I variant E16N.[22] The chromatograms shown belong to an extracted ion current for the singly (1905.9935) and doubly (953.5006) charged ions of the lasso peptide before and after treatment at 95 °C for 1 h. The peaks are labeled as branched-cyclic or lasso peptide, respectively, and the observed masses for the doubly protonated species are shown adjacent to them.

the only factor affecting the general stability of a lasso peptide, even when identically sized macrolactam rings are present. Therefore, future studies will focus on both unveiling other criteria important for the maintenance of the lasso fold and on characterizing the range of amino acids that are suitable for stabilizing a lasso peptide with a nine-residue macrolactam ring.

4.1.3 The Allowed Sizes of the Macrolactam Rings in Lasso Peptides

Since the isolation of anantin until very recently, the only macrolactam rings found in lasso peptides were eight- and nine-residues in size.[10,17,18,22-25,27-30,32,34,37,40,41,43-46,48,126] Thus, it is of little surprise that this observation led to the idea that these are the only ring sizes possible for lasso peptides. The fundamental argumentation for this assumption was that ten-residue macrolactam rings would be too big to allow the steric maintenance of the lasso fold, while the opening in a seven-residue macrolactam ring was deemed to be too small to allow the threading by the *C*-terminal tail. The latter assumption was proven wrong by the report of the xanthomonins I-III, which are extremely stable lasso peptides with seven-residue macrolactam rings.[26] Nevertheless, the first indications for the possibility of seven-residue macrolactam rings in lasso peptides were already found in an earlier genome mining study.[37] In this investigation, several instances became apparent, where putative lasso peptide sequences contained a glutamate residue suitable for ring formation at position 7. In some cases, like xanthomonins I-II, an additional glutamate or aspartate residue was present at positions 8 or 9 of the predicted lasso peptide, while in the remaining instances, like xanthomonin III, only the position 7 residue of the putative lasso sequence featured a carboxylic acid side chain. Indeed, during the first investigations of the heterologous expression of the gene cluster producing xanthomonins I and II, we still expected the production of ordinary eight-residue

macrolactam ring containing lasso peptides. Conversely, the first tandem MS analyses of these compounds already included strong hints for the presence of seven-residue macrolactam rings, as the fragmentation spectra showed strong signals corresponding to the first seven amino acids minus the one molecule of water that is lost during the cyclization reaction (Figure 4.4).[26]

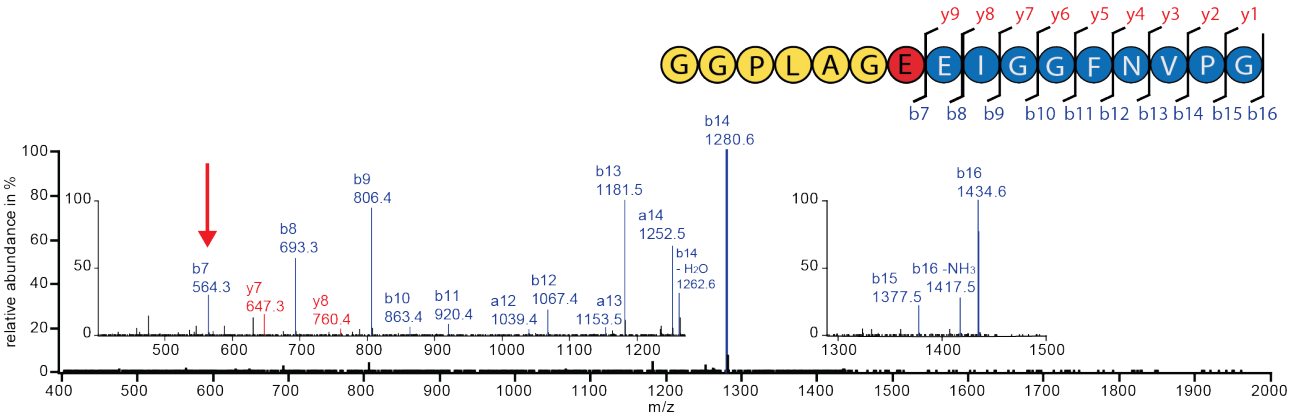


Figure 4.4. Fragmentation spectrum of xanthomonin I.[26] The b- and y-fragments are depicted in blue or red, respectively. The b7 fragment is furthermore highlighted by a red arrow.

Subsequent generation and expression of E7A and E8A mutants of xanthomonin II already conclusively proved the ring size predicted by MS, as the E7A variant was not produced, while the E8A variant was produced on wild type level.[26] Finally, the structure elucidations of both xanthomonin I and II revealed the spacial organizations of these small rings. Due to these structures, it also became apparent how small the ring openings really were and that they could barely be threaded by the *C*-terminal tails.[26] This observation was further substantiated by results from the mutational studies on xanthomonin II, which confirmed that the only amino acid tolerated at the two positions that thread the ring was glycine and that even the introduction of a barely larger alanine residue at these positions completely or almost completely abolished the lasso peptide production.[26] Combining this information with what is known for the chemically accessible class of the [2]rotaxanes,[127-129] it becomes likely that seven-residues compose the lower limit possible for threaded macrolactam rings. In regard to ten-residue macrolactam rings, research has yet to prove or disprove their occurrence in nature. Still, so far no hints for lasso peptides with such an arrangement were found, but then again, this held true for lasso peptides featuring seven-residue rings until it was recently proven otherwise.

4.1.4 The *N*-terminal Amino Acids of Lasso Peptides

Similarly to the presence of seven-residue macrolactam rings, the presence of an *N*-terminal amino acid different to cysteine or glycine was assumed unlikely due to the lack of previous observations.[10,17,18,22-30,32,34,37,40,41,43-46,48,126] Nonetheless, several putative class II lasso peptide precursors were recently identified that featured the conserved threonine residue at the penultimate position of the leader peptide, but lacked the usual glycine residue at position 1 of the lasso sequence. In some of these instances, a glycine was present one amino acid closer to the *C*-terminus.[37] These precursors rose the question, if they were really matured into a lasso peptide featuring a so far unobserved amino acid at their first position or if they indeed contained a change of the the TxG cleavage motif to TxxG (Figure 4.5).

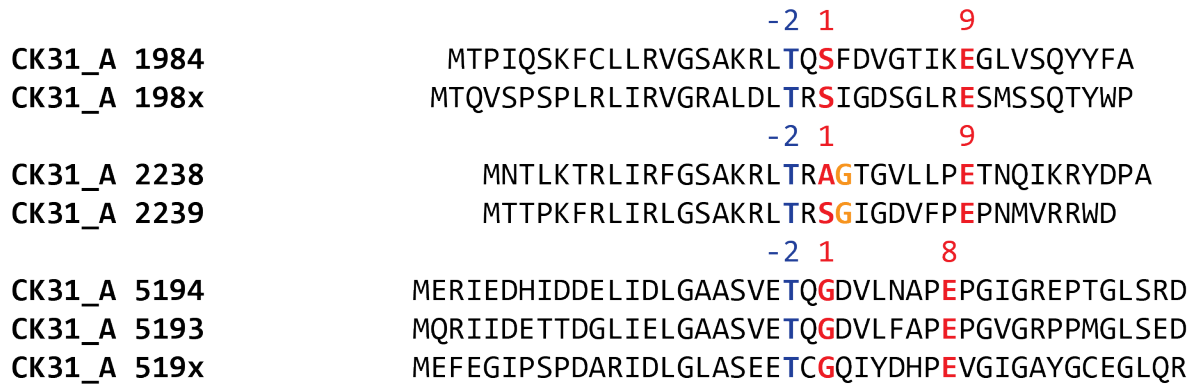


Figure 4.5. Schematic depiction of the precursor peptides of caulonodins I-VII.[37,47] While the precursor peptides of caulonodins I-III (CK31_A 5194/5193/519x) fulfill all previously postulated criteria of lasso peptide precursors, the precursor peptides of caulonodins IV/V (CK31_A 1984/198x) and caulonodins VI/VII (CK31_A 2238/2239) feature an unusual amino acid at position 1 of the lasso sequence. The T-2 residue is highlighted in blue, the position 1 residues and ring forming glutamates in red, while all glycines at position 2 of a lasso sequence are shown in orange.

Additionally, it could not be excluded that in these instances a simple random mutation event had occurred, altering the glycine at position 1 to a different amino acid and thus abolishing lasso peptide production in the corresponding organism. Due to the availability of genetic tools and the gathered experience for significant improvement of heterologous production of proteobacterial lasso peptides in *E. coli*, it was decided upon investigation of two gene clusters from *Caulobacter* sp. K31 that contained four precursor peptides with this peculiar feature.[47] The subsequent experiments not only proved that these lasso peptides could be produced in good yields in *E. coli*, but also showed that a glycine exchange of the serine or alanine residues at position 1 of these lasso peptides drastically decreased the overall lasso peptide production.[47] These results are converse to previously reported mutational studies for lasso peptides carrying a glycine at position 1. In these

cases, it was shown that the glycine could not or only poorly be exchanged to any other amino acid (Figure 4.6).[22,25,36,62,67] Taken together, these different studies show that lasso peptides do not appear to be generally restricted to carry either a glycine or cysteine residue at position 1, but at the same time also reveal that the present position 1 amino acid seems to be recognized with high specificity by its corresponding biosynthetic machinery. Thus, in highlight of these results, the scope of small open reading frames that could encode class II lasso peptide precursors was extended to theoretically allow every canonical amino acid at position 1 of the lasso sequence, even though the predominant amino acid at this position still appears to be glycine.

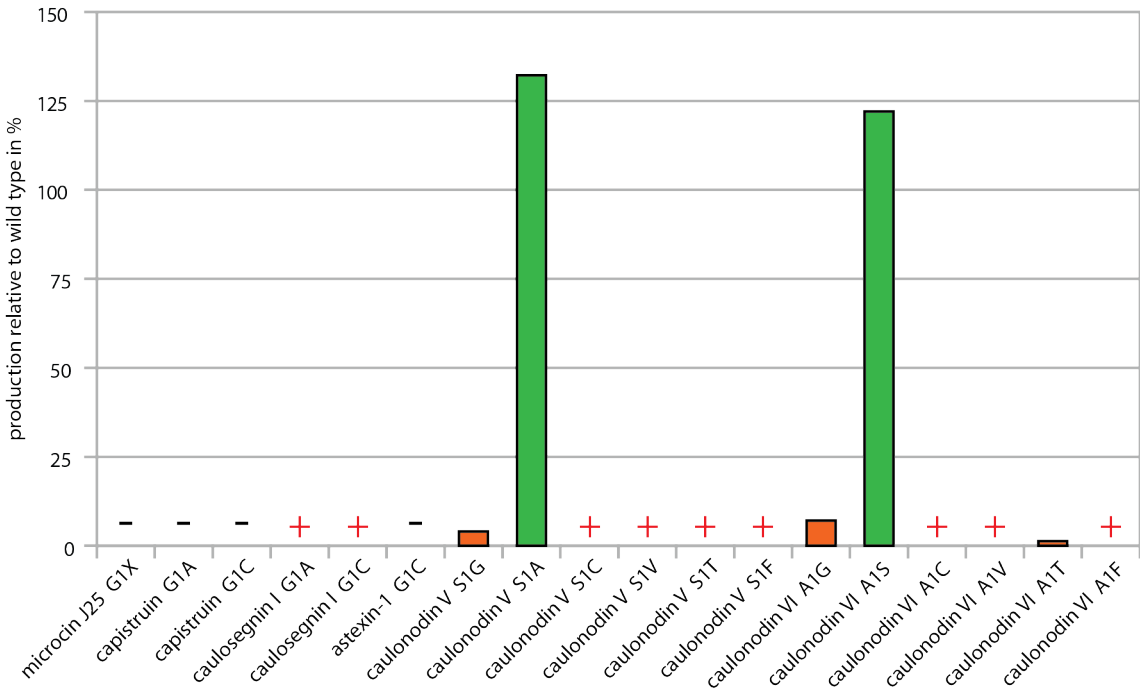


Figure 4.6. Production of all reported position 1 lasso peptide variants.[22,25,36,47,62] The bars are highlighted in different colors corresponding to the observed relative production. A green bar represents a production of >50%, a yellow bar a production of 50-15% and an orange bar a production of 15-1% relative to the respective wild type production levels. Variants that were only detectable by MS are marked with a red cross, while variants that showed no production at all are highlighted with a black minus sign. In case of microcin J25, the G1X refers to this residue being exchanged against all other canoncial amino acids of which non of the corresponding variants showed any production.

4.1.5 The Threonine at the Penultimate Position of the Leader Peptide as Essential Residue in Lasso Peptide Precursors

The importance of the threonine residue at the penultimate position of the leader sequence for lasso peptide maturation was first recognized for the MccJ25 and the capistruin systems.[63-65] Since then, closer inspection of this residue revealed that its exchange against another amino acid practically always negatively affects the overall lasso peptide production. This effect is weakest, when the

newly introduced amino acid shares some structural similarity to threonine, which is the case for serine, cysteine, valine and isoleucine.[64] In subsequent mutational studies with different lasso peptide processing systems, it was furthermore shown that the tolerance for such structurally conserved mutations of T-2 differs from system to system.[22,25,26] Apparently, some processing enzymes, like the ones that transform McjA to MccJ25, tolerate these alterations rather well, while other systems, like the one maturing xanthomonins I and II, react very poorly to such a substitution with regard to the overall efficiency of lasso peptide production (Figure 4.7).[22,25,26,63-65]

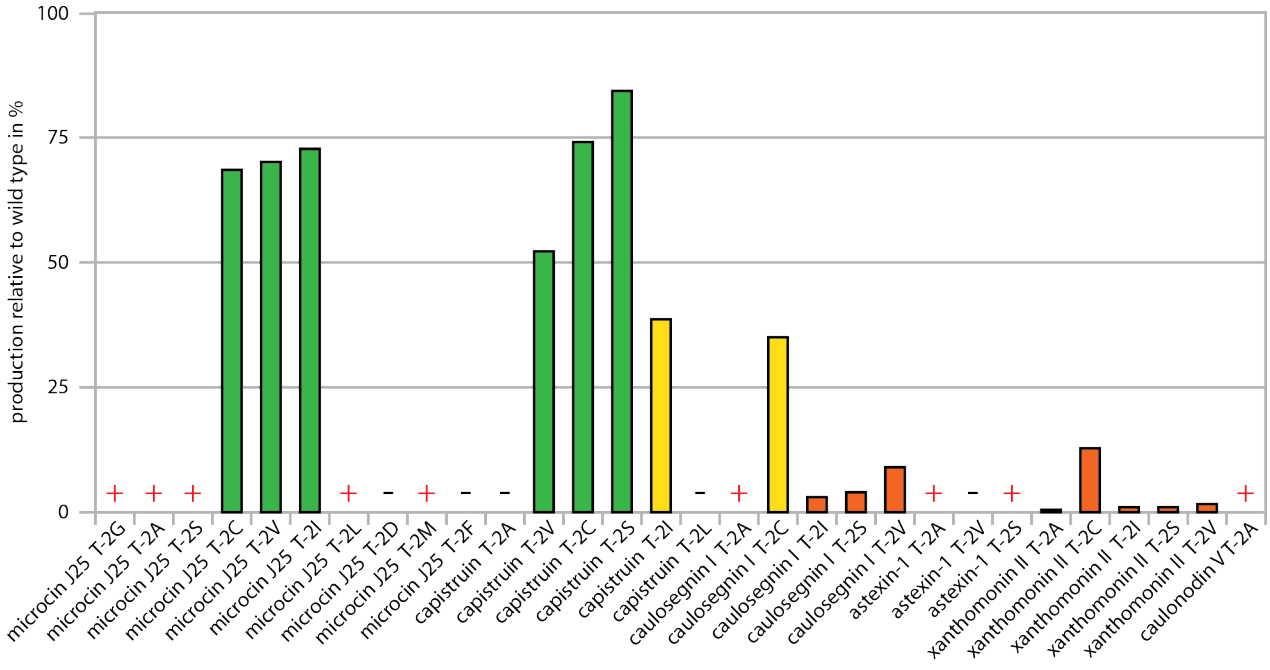


Figure 4.7. Lasso peptide production levels that correspond to all assayed T-2 lasso peptide precursor variants reported in the literature.[22,25,26,47,64] The bars are highlighted in different colors corresponding to the observed relative production. A green bar represents a production of >50%, a yellow bar a production of 50-15% and an orange bar a production of 15-1% relative to the respective wild type production levels. Variants that were only detectable by MS are marked with a red cross, while variants that showed no production at all are highlighted with a black minus sign.

Thus, the recognition of the T-2 residue seems to be of an universal nature and to be conserved for all lasso peptide processing enzymes, even though some show at least a basic tolerance towards changes. This observation holds true until today with no known example of a functional, natural occurring lasso peptide precursor that does not feature a threonine at this position. Still, on basis of the first McjA precursor peptide studies, it was assumed that T-2 is the only strictly conserved leader peptide residue that is crucial for enzymatic recognition.[63] This assumption turned out to be at least partially wrong, as we recently reported a generally conserved recognition motif in the *C*-terminal section of a lasso peptide leader sequence.[47] The reason for calling this assumption only

partially and not generally wrong is that this motif seems to be only strictly conserved for the precursor peptides encoded by ABC-type proteobacterial lasso peptide biosynthetic gene clusters, which explains why it is not present in the MccJ25 or capistruin precursors. Nonetheless, this finding suggests that corresponding motifs may be found in one form or the other in each of the different subgroups of lasso peptide biosynthetic gene cluster arrangements. At the same time, it also shows that even though a large number of the conserved positions and residues may vary in between these different subgroups, the threonine at the penultimate position of the leader peptide seems to be conserved in all or at least most of them. Therefore, the notion that only T-2 is essential for enzymatic recognition appears to be flawed and should be considered accordingly in future mutational studies of and genome mining searches for novel lasso peptides.

4.1.6 The Antimicrobial Activity of Lasso Peptides

There are several instances known, where lasso peptides show a significant antimicrobial activity against certain bacteria[10,17,18,28,32,34,41,43] and interestingly most of these compounds were identified in studies that were not focused on the discovery of a new lasso peptide itself. The only exceptions to this are capistruin[18] and maybe astexin-1,[30] even though the latter exhibits such a weak antimicrobial activity against *Caulobacter crescentus* that it is unclear if this activity is of any real relevance. Because of the occurrence of an antimicrobial activity of such lasso peptides in combination with the observation that MccJ25 is only produced by its host under nutrient-limitation, it was reasoned that lasso peptides generally could be antimicrobial agents that ensure their hosts survival under stressful environmental conditions. Hence, it was surprising that a large number of recently isolated lasso peptides that were tested for such an activity, appeared to be completely inactive in this regard. Although a more thorough screening might reveal sensitive organisms, it may be that these compounds just generally do not exhibit an antimicrobial activity. Interestingly, all known biosynthetic gene clusters producing antimicrobially active lasso peptides, with exception of the questionably active astexin-1, contain a gene encoding an ATP-dependent transporter (Figure 4.8), while gene clusters that produce supposedly non-antimicrobial lasso peptides lack such a gene.

Due to the low number of known examples in both cases, this observation might just be coincidental, but from a logical point of view it makes sense to reason that expression of a dedicated ABC-transporter is only necessary in presence of an antimicrobially active compound. In conclusion, this new data puts the assumption of a general role of lasso peptides as antimicrobial agents in question and suggests that this is more likely only true for a subgroup of this natural product family. The implications of this observation is discussed in more detail in the following section.

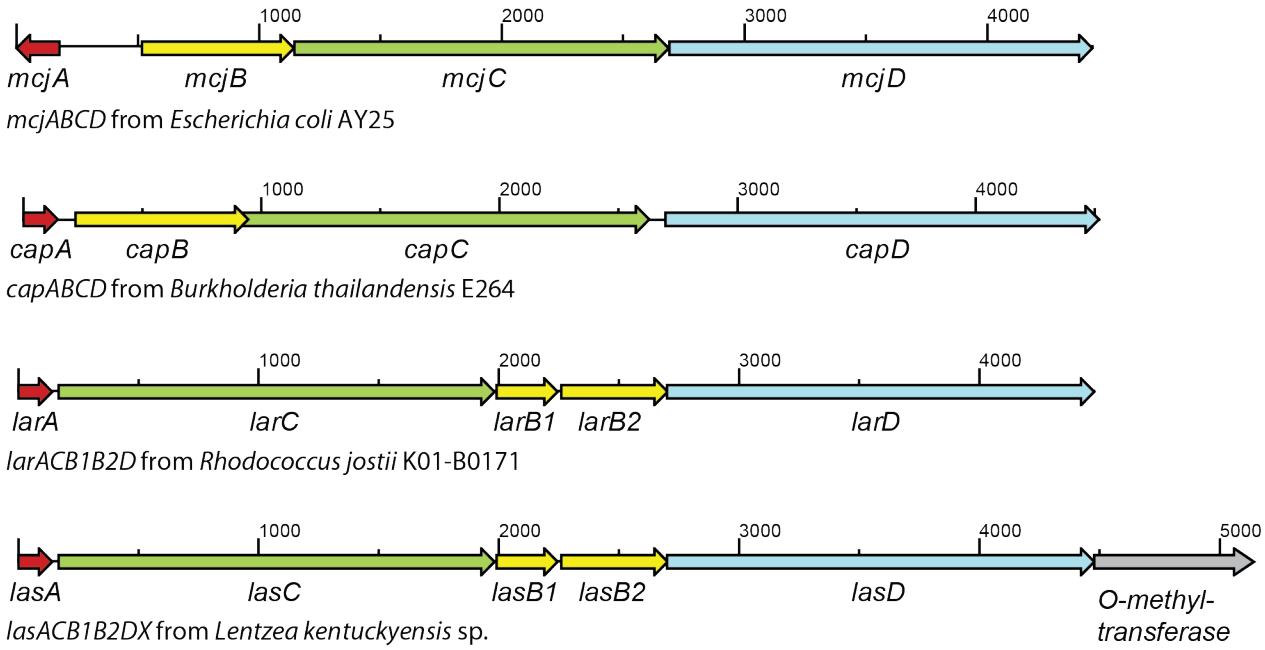


Figure 4.8. Schematic representation of all known biosynthetic gene clusters that produce highly active antimicrobial lasso peptides.[18,28,55,57,58] The names of the genes from the lariatin and lassomycin biosynthesis systems were altered to fit the established nomenclature for lasso peptide biosynthetic gene clusters according to the previously described reasoning.[37]

4.2 New Genome Mining Studies Hint at the Possible Functions of Lasso Peptides in Nature

Before 2008, lasso peptides were but a mere curiosity in the field of natural product research. Their discovery was left to chance and limited to the few examples where they were both strongly produced by their hosts and coincidentally also conferred a bioactivity, which was looked for by the researchers.[10,17,27,32,34,40,41,43-46,48] This all changed with the report of capistruin, the first lasso peptide found by genome mining efforts intended to actually discover a new representative of this natural product family.[18] Since then, the coincidental isolation of novel lasso peptides became scarce[24,28,29] and was more and more replaced by directed, systematic approaches with the aim to isolate new examples of as well as to learn more about this unique class of RiPPs.[18,22,23,26,30,37] Through these efforts, it became clear that lasso peptides are distributed amongst almost all bacterial phyla and in some cases even constitute one of the only few secondary metabolites produced by a particular organism.[22,30,37,47] Nevertheless, the overall role of these compounds in nature remains enigmatic. With some instances known where they can act as antimicrobials[10,17,18,27,28,32,34,40,41,43] and others were they confer an interesting, but most likely off-target activity,[10,24,27,34,40,41,43-46,48] the assumption became prevalent that these compounds were generally produced to kill off competitors and ensure survival

of their hosts under stressful environmental conditions. The general truth of this notion was dispelled as more and more examples of lasso peptides were reported that lack any antimicrobial activity.[22,23,26,37] Of course, it cannot be ruled out that in these instances merely the wrong organisms were tested for sensitivity, but still it seems likely that at least some lasso peptides could have a completely different function. The idea of no general role for lasso peptides that is true independent from where they originated is underlined by the observation that the corresponding biosynthetic gene clusters can be separated into different subgroups, depending on the respective cluster organization and their genetic surroundings.[23,30,37] Thus, it may be possible that for example the antimicrobial fraction of lasso peptides always occurs together with dedicated immunity conferring enzymes like McjD, while the lack of such an enzyme could be a direct proof of a lasso peptide not exhibiting a bactericidal activity. Of course, at this point it is completely unclear if only ABC-transporters can fulfill the role as lasso peptide immunity factors. On the contrary, it would be rather peculiar if different clades of bacteria with different kinds of lasso peptide biosynthetic gene clusters would not at least in some cases have developed alternative mechanisms to ensure survival of the host organism when it is producing an antimicrobial active lasso peptide. Still, ABC-transporters remain the only known examples for this so far, highlighting that the eyes of the researchers should be kept pried open for alternative ways and mechanisms for bacteria to deal with this subclass of bioactive lasso peptides. In this regard, it should be mentioned that there is at least one known instance where an organism encodes enzymes which can specifically degrade the lasso peptides it produces. These enzymes are the isopeptidases produced by *Asticcacaulis extrenticus* CB48, which were shown to cleave the isopeptide bonds of lasso peptides produced by this organism and in this way transform them back into linear peptides that are easily degraded by other proteases. Furthermore, these enzymes are highly specific, i.e. they only recognize peptides in a lasso fold and not their branched-cyclic analogues and each isopeptidase just degrades the lasso peptides produced by one single biosynthetic machinery.[23] Interestingly, the genes encoding these isopeptidases are part of a gene cluster adjacent to the respective biosynthetic gene clusters that produce the recognized lasso peptides. Genome mining studies furthermore revealed that the co-occurrence of an ABC-type lasso peptide biosynthetic gene cluster and an adjacent isopeptidase encoding gene cluster seems to be highly conserved throughout proteobacteria.[23,37] Still, it remains doubtful that these isopeptidases are indeed immunity factors. The reason for this notion is the lack of a significant antimicrobial activity of any of the newly isolated lasso peptides that were tested in this regard. In addition to this observation, a closer inspection of the other genes next to the ones encoding the respective isopeptidases also suggests a different function for these systems (Figure 4.9).

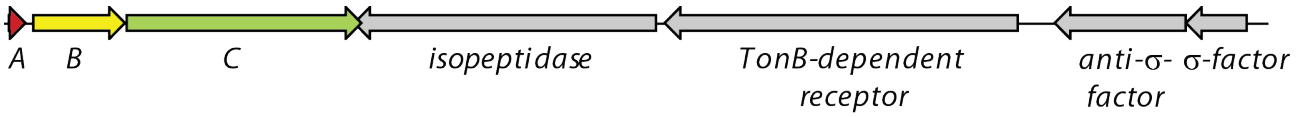


Figure 4.9. Schematic depiction of the basic organization of an ABC-type lasso peptide biosynthetic gene cluster and an adjacent isopeptidase gene containing cluster.[37]

Normally, the first two genes in these clusters encode a putative σ-factor and an anti-σ-factor, respectively. The next gene then usually encodes a putative TonB-dependent receptor and the fourth and last gene of these kind of cluster arrangements encodes the actual isopeptidase. Especially the presence of a TonB-dependent receptor, a protein that practically has the exact opposite function to an ATP-dependent exporter, suggests that the gene products of these clusters more likely play a role in sensing, uptake and subsequent degradation of the corresponding lasso peptides instead of conferring a resistance against supposedly antimicrobially active compounds. In this context, the question arises, why an organism should selectively take up a compound that it also just could produce itself. The most likely answer for this would be that these molecules either act as signaling or as scavenger molecules. To resolve this, future studies need to address the function and role of the complete isopeptidase gene clusters, either by heterologously producing and characterizing the remaining three gene products adjacent to the isopeptidase gene or by incorporating mutations in the corresponding genomes with subsequent probing what differences occur if the complete or at least parts of such a cluster are knocked-out. In addition to this, supplemental studies should be performed on clusters that differ from both the well-known ABC- or ABCD-type arrangements, as this could reveal yet another subclass of lasso peptides that could possibly also have another biological function.

4.3 Potential of Modified Lasso Peptides for Medicinal Applications

Prior to the experiments presented here, a proof-of-concept study was published, detailing that it is generally possible to incorporate a bioactive peptide epitope into a lasso scaffold and thus creating a hybrid compound with both the bioactivity conferred by the peptidic pharmacophore and the physiological stability of the lasso fold.[65] Even though this study demonstrated the general possibility of such an endeavor, the resulting lasso graft had only an unsatisfactory binding affinity and selectivity for its target. Thus, we wanted to assess, if the basic bioactivity this graft exhibited could be improved. For this, a rational approach based on previous studies that were involved in the development of the peptidic drug cilengitide was chosen. Indeed, both consequently generated lasso peptide variants had an improved binding selectivity and in one case additionally a higher binding

affinity towards the target αvβ3 integrin receptor. The subsequent NMR based structure elucidation of MccJ25(RGDF) and molecular dynamics simulations with this structure furthermore confirmed that the improved binding behavior most likely relates to direct interactions of the newly incorporated phenylalanine with the receptor binding pocket.[69] Taken together, these results demonstrate that grafted lasso peptides can indeed be optimized by rational approaches. Nonetheless, this study also revealed some still existing drawbacks of the use of lasso peptides in epitope grafting efforts. Besides the obvious limitation to canonical amino acids due to the persisting lack of a chemical route to lasso peptide synthesis, the incorporation of the desired mutations in the lasso peptide itself can already prove to be problematic. For example, although a manifold of unpublished mutants was generated in the course of the work on the improvement of the MccJ25(RGD) graft and even though their maturation was confirmed by LC-MS analysis, the alterations introduced in these variants reduced the lasso peptide yields so much that their isolation was not possible. In these attempts, the improved binding behavior should be achieved by either switching the position of the peptide epitope or by increasing the structural rigidity of the binding motif by incorporation of two cysteines that were able to form a disulfide bond. Thus, even though these variants were generated on a sound reasoning, the unsuccessful isolation of these compounds prohibited the assessment of them in regard to their αvβ3 binding behavior. Therefore, the introduction of a phenylalanine adjacent to the RGD binding motif was chosen, which finally not only yielded lasso grafts that were produced in amounts suitable for isolation, but which at the same time also showed the desired improved binding behavior. So, even though a positive result was obtained in the end of this project, in future studies the aforementioned problems could either completely prevent the incorporation of other peptide epitopes or at least prohibit the further improvement of an established, active lasso graft. Additionally, this problem also causes a challenge in regard to the high-throughput generation of diverse compound libraries for activity assessments. Although a general approach for this was shown in a study where randomly generated MccJ25 variants were screened for an improved bactericidal activity, this was only possible due to the fact that the target activity allowed a very straightforward live-and-death-screening of the newly generated variants.[66] Thus, every other target activity would require the prior development and establishment of another, more sophisticated system that allows a screening in this or a similar way, which is not always possible. In conclusion, lasso peptides generally do show potential in the field of epitope grafting approaches and for the development of drug leads. Nevertheless, the results obtained so far, though promising, emphasize that more work has to be done to increase the tool set to facilitate both the screening and development of novel bioactive lasso grafts and thus allowing an access to more efficient and more extensive ways to produce compound diversity.

5. Outlook

5.1 The Next Obstacles in Lasso Peptide Research

Lasso peptide research has made big advances in recent years. Not only has the number of known lasso peptides and corresponding biosynthetic gene clusters drastically increased, but also more and more information about the biosynthesis, physico-chemical properties and distribution in nature of these interesting compounds has come to light. Nevertheless, since the isolation of the first lasso peptides many questions still remain unresolved, which have to be addressed in future research. The new research tasks derived from these questions include, but are not limited to, the search for novel and so far uncharacterized types of lasso peptide biosynthetic gene clusters, getting a deeper understanding of the maturation enzymes including the underlying mechanism of the lasso peptide biosynthesis itself, studying their physico-chemical behaviors as well as finding out the reasons for differences of such properties for different lasso peptides and finding out more about their role in nature. For this, we can apply the lessons we already learned about lasso peptides and how to produce them, but we will also have to face new challenges when moving away from the familiar proteobacterial systems and have to start to find new solutions for upcoming problems.

One of the biggest problems in this regard will be the incompatibility of some of the non-proteobacterial biosynthetic gene clusters with *E. coli* expression. This may be caused by obvious reasons like a different codon bias, but can also include more complex problems like the processing enzymes being adapted to a biochemical environment different to the one present inside *E. coli* cells. In these cases, several approaches could be pursued and future studies have to reveal which of them are most efficient for obtaining positive results. These approaches range from the *de novo* gene synthesis after optimization of the nucleotide sequence in respect to *E. coli* codon usage, over the complete change of the heterologous expression system to the thorough screening of a multitude of homologous fermentation conditions to find conditions favoring the lasso peptide production by the original host organism. At least in the case of switching the heterologous host, there is one report where this was already done successfully. In this study, the biosynthetic gene cluster responsible for the lariatin production was identified in the genome of *Rhodococcus jostii* K01-B0171, cloned in a suitable vector system and finally heterologously produced in a lariatin deficient Rhodococcus mutant strain.[57] Of course, in this case the heterologous host was almost completely identical with the producing strain and as the authors did not comment on the yields of their heterologous lasso peptide production, it most likely was not too impressive, but still this study can be seen as a proof-of-principle that encourages the application of a similar methodology for future research. Such approaches are nowadays furthermore facilitated by the fact that more and

more genetically engineered subspecies of organisms like *Bacillus subtilis*,[131] *Streptomyces coelicolor*[132] or *Lactococcus lactis*[133] are established as heterologous expression hosts. These systems additionally become increasingly attractive due to the steady optimization of corresponding genetic tools, which leads to the continuous improvement of the overall expression efficiency in these organisms. Still, even though it most likely would be beneficial to change the expression system from *E. coli* to a host that is more closely related to the strain the biosynthetic gene cluster in question originated from, this could also necessitate the development of a completely new optimization approach than the one we successfully employed for the proteobacterial lasso peptide biosynthetic gene clusters. Nevertheless, once established, such a new approach could hopefully be applied to all or at least most of the clusters found in bacteria belonging to the corresponding phylum. As such, the development of this would be greatly desirable as it then could provide access to a so far mostly untapped source for lasso peptides, in particular when they are derived from clusters with more unusual genetic arrangements. In this regard, biosynthesis systems like the ones involved in the production of the RES-701 family lasso peptides or lassomycin seem to be especially interesting, as they may provide information about lasso peptide tailoring enzymes, which is something research utterly neglected so far.

Still, future studies should not solely focus on the isolation of new lasso peptides, but also try to unveil more information about the mechanism of lasso peptide biosynthesis in general. So far, the research in this area is proceeding very slowly as the lasso peptide maturation enzymes are generally very challenging to work with. This notion is emphasized by all the unsuccessful attempts of getting soluble, active proteins from a manifold of different clusters that were made in all the years where lasso peptide research was conducted in this group. Then again, this general problem only means that a success in this regard would be even more interesting, as it may finally allow access to not only a more detailed characterization of the processing enzymes, but maybe also could lead to structural investigations of these enzymes and their interactions with the precursor peptides by means of X-ray crystallography, H-D exchange experiments and the likes.

Finally, future research should strive to establish the role of lasso peptides in nature. For this, current leads like the proteobacterial lasso peptide isopeptidase containing gene clusters as well as the systems that produce antimicrobial lasso peptides should be primarily pursued, but at the same time completely new and unknown systems should not be completely ignored either.

In summary, even though plenty of new information about lasso peptides was obtained in recent years, many more insights about this intriguing family of natural products are just waiting to be uncovered.

5.2 Epitope Grafting of Lasso Peptides

Since it was both shown to be possible to successfully graft peptide epitopes into lasso peptide scaffolds[65] as well as to furthermore improve these bioactive grafts by rational means,[69] the next goal should be to apply these techniques to new epitopes and new biological targets. For this, all peptide epitopes that use canonical amino acids and have a size of up to five residues seem to be promising starting points at the moment. Still, for this research to excel, additional studies assessing the use of lasso peptide scaffolds different to MccJ25 as well as testing ways to utilize methods for the *in vivo* incorporation of non-canonical amino acids and the chemical modification of isolated lasso peptides should be conducted simultaneously. These complementary studies could help to emphasize the potential of lasso peptides as grafting scaffolds and could facilitate an efficient access to new and more diverse lasso peptide variants.

As such, it becomes clear that epitope grafting with lasso peptides is still in its infancy and currently cannot rival already well-established grafting platforms like cyclotides. Nonetheless, some characteristics of lasso peptides would nicely complement these well-known systems and thus it seems promising to further pursue the use of lasso peptides in drug development.

6. References

[1] Giessen, T. W.; Marahiel, M. A.; *Ribosome-independent biosynthesis of biologically active peptides: Application of synthetic biology to generate structural diversity*; *FEBS Lett* **2012**, *586*, 2065-2075.

[2] Arnison, P. G.; Bibb, M. J.; Bierbaum, G.; Bowers, A. A.; Bugni, T. S.; Bulaj, G.; Camarero, J. A.; Campopiano, D. J.; Challis, G. L.; Clardy, J.; Cotter, P. D.; Craik, D. J.; Dawson, M.; Dittmann, E.; Donadio, S.; Dorrestein, P. C.; Entian, K. D.; Fischbach, M. A.; Garavelli, J. S.; Goransson, U.; Gruber, C. W.; Haft, D. H.; Hemscheidt, T. K.; Hertweck, C.; Hill, C.; Horswill, A. R.; Jaspars, M.; Kelly, W. L.; Klinman, J. P.; Kuipers, O. P.; Link, A. J.; Liu, W.; Marahiel, M. A.; Mitchell, D. A.; Moll, G. N.; Moore, B. S.; Muller, R.; Nair, S. K.; Nes, I. F.; Norris, G. E.; Olivera, B. M.; Onaka, H.; Patchett, M. L.; Piel, J.; Reaney, M. J.; Rebuffat, S.; Ross, R. P.; Sahl, H. G.; Schmidt, E. W.; Selsted, M. E.; Severinov, K.; Shen, B.; Sivonen, K.; Smith, L.; Stein, T.; Sussmuth, R. D.; Tagg, J. R.; Tang, G. L.; Truman, A. W.; Vederas, J. C.; Walsh, C. T.; Walton, J. D.; Wenzel, S. C.; Willey, J. M.; van der Donk, W. A.; *Ribosomally synthesized and post-translationally modified peptide natural products: overview and recommendations for a universal nomenclature*; *Nat Prod Rep* **2012**, *30*, 108-160.

[3] Toyama, H.; Fukumoto, H.; Saeki, M.; Matsushita, K.; Adachi, O.; Lidstrom, M. E.; *PqqC/D, which converts a biosynthetic intermediate to pyrroloquinoline quinone*; *Biochem Biophys Res Commun* **2002**, *299*, 268-272.

[4] Tsai, T. Y.; Yang, C. Y.; Shih, H. L.; Wang, A. H.; Chou, S. H.; *Xanthomonas campestris PqqD in the pyrroloquinoline quinone biosynthesis operon adopts a novel saddle-like fold that possibly serves as a PQQ carrier*; *Proteins* **2009**, *76*, 1042-1048.

[5] Wecksler, S. R.; Stoll, S.; Iavarone, A. T.; Imsand, E. M.; Tran, H.; Britt, R. D.; Klinman, J. P.; *Interaction of PqqE and PqqD in the pyrroloquinoline quinone (PQQ) biosynthetic pathway links PqqD to the radical SAM superfamily*; *Chem Commun (Camb)* **2010**, *46*, 7031-7033.

[6] Freeman, M. F.; Gurgui, C.; Helf, M. J.; Morinaka, B. I.; Uria, A. R.; Oldham, N. J.; Sahl, H. G.; Matsunaga, S.; Piel, J.; *Metagenome mining reveals polytheonamides as posttranslationally modified ribosomal peptides*; *Science* **2012**, *338*, 387-390.

[7] Shen, Y. Q.; Bonnot, F.; Imsand, E. M.; RoseFigura, J. M.; Sjolander, K.; Klinman, J. P.; *Distribution and properties of the genes encoding the biosynthesis of the bacterial cofactor, pyrroloquinoline quinone*; *Biochemistry* **2012**, *51*, 2265-2275.

[8] Tianero, M. D.; Donia, M. S.; Young, T. S.; Schultz, P. G.; Schmidt, E. W.; *Ribosomal route to*

small-molecule diversity; *J Am Chem Soc* **2012**, *134*, 418-425.

[9] Frechet, D.; Guitton, J. D.; Herman, F.; Faucher, D.; Helynck, G.; Monegier du Sorbier, B.; Ridoux, J. P.; James-Surcouf, E.; Vuilhorgne, M.; *Solution structure of RP 71955, a new 21 amino acid tricyclic peptide active against HIV-1 virus*; *Biochemistry* **1994**, *33*, 42-50.

[10] Potterat, O.; Stephan, H.; Metzger, J. W.; Gnau, V.; Zähner, H.; Jung, G.; *Aborycin - a tryciclic 21-peptide antibiotic isolated from Streptomyces griseoflavus*; *Liebigs Ann. Chem.* **1994**, 741-743.

[11] Constantine, K. L.; Friedrichs, M. S.; Detlefsen, D.; Nishio, M.; Tsunakawa, M.; Furumai, T.; Ohkuma, H.; Oki, T.; Hill, S.; Bruccoleri, R. E.; et al.; *High-resolution solution structure of siamycin II: novel amphipathic character of a 21-residue peptide that inhibits HIV fusion*; *J Biomol NMR* **1995**, *5*, 271-286.

[12] Detlefsen, D. J.; Hill, S. E.; Volk, K. J.; Klohr, S. E.; Tsunakawa, M.; Furumai, T.; Lin, P. F.; Nishio, M.; Kawano, K.; Oki, T.; et al.; *Siamycins I and II, new anti-HIV-1 peptides: II. Sequence analysis and structure determination of siamycin I*; *J Antibiot (Tokyo)* **1995**, *48*, 1515-1517.

[13] Katahira, R.; Shibata, K.; Yamasaki, M.; Matsuda, Y.; Yoshida, M.; *Solution structure of endothelin B receptor selective antagonist RES-701-1 determined by 1H NMR spectroscopy*; *Bioorg Med Chem* **1995**, *3*, 1273-1280.

[14] Bayro, M. J.; Mukhopadhyay, J.; Swapna, G. V.; Huang, J. Y.; Ma, L. C.; Sineva, E.; Dawson, P. E.; Montelione, G. T.; Ebright, R. H.; *Structure of antibacterial peptide microcin J25: a 21-residue lariat protoknot*; *J Am Chem Soc* **2003**, *125*, 12382-12383.

[15] Rosengren, K. J.; Clark, R. J.; Daly, N. L.; Goransson, U.; Jones, A.; Craik, D. J.; *Microcin J25 has a threaded sidechain-to-backbone ring structure and not a head-to-tail cyclized backbone*; *J Am Chem Soc* **2003**, *125*, 12464-12474.

[16] Wilson, K. A.; Kalkum, M.; Ottesen, J.; Yuzenkova, J.; Chait, B. T.; Landick, R.; Muir, T.; Severinov, K.; Darst, S. A.; *Structure of microcin J25, a peptide inhibitor of bacterial RNA polymerase, is a lassoed tail*; *J Am Chem Soc* **2003**, *125*, 12475-12483.

[17] Iwatsuki, M.; Tomoda, H.; Uchida, R.; Gouda, H.; Hirono, S.; Omura, S.; *Lariatins, antimycobacterial peptides produced by Rhodococcus sp. K01-B0171, have a lasso structure*; *J Am Chem Soc* **2006**, *128*, 7486-7491.

[18] Knappe, T. A.; Linne, U.; Zirah, S.; Rebuffat, S.; Xie, X.; Marahiel, M. A.; *Isolation and structural characterization of capistruin, a lasso peptide predicted from the genome sequence of Burkholderia thailandensis E264*; *J Am Chem Soc* **2008**, *130*, 11446-11454.

[19] Knappe, T. A.; Linne, U.; Xie, X.; Marahiel, M. A.; *The glucagon receptor antagonist BI-32169 constitutes a new class of lasso peptides*; *FEBS Lett* **2010**, *584*, 785-789.

[20] Nar, H.; Schmid, A.; Puder, C.; Potterat, O.; *High-resolution crystal structure of a lasso*

Peptide; *ChemMedChem* **2010**, *5*, 1689-1692.

[21] Xie, X.; Marahiel, M. A.; *NMR as an effective tool for the structure determination of lasso peptides*; *Chembiochem* **2012**, *13*, 621-625.

[22] Hegemann, J. D.; Zimmermann, M.; Xie, X.; Marahiel, M. A.; *Caulosegnins I-III: a highly diverse group of lasso peptides derived from a single biosynthetic gene cluster*; *J Am Chem Soc* **2013**, *135*, 210-222.

[23] Maksimov, M. O.; Link, A. J.; *Discovery and characterization of an isopeptidase that linearizes lasso peptides*; *J Am Chem Soc* **2013**, *135*, 12038-12047.

[24] Um, S.; Kim, Y. J.; Kwon, H.; Wen, H.; Kim, S. H.; Kwon, H. C.; Park, S.; Shin, J.; Oh, D. C.; *Sungsanpin, a lasso peptide from a deep-sea streptomycete*; *J Nat Prod* **2013**, *76*, 873-879.

[25] Zimmermann, M.; Hegemann, J. D.; Xie, X.; Marahiel, M. A.; *The astexin-1 lasso peptides: biosynthesis, stability, and structural studies*; *Chem Biol* **2013**, *20*, 558-569.

[26] Hegemann, J. D.; Zimmermann, M.; Zhu, S.; Steuber, H.; Harms, K.; Xie, X.; Marahiel, M. A.; *Xanthomonins I–III: A New Class of Lasso Peptides with a Seven-Residue Macrolactam Ring*; *Angewandte Chemie International Edition* **2014**, *53*, 2230-2234.

[27] Chokekijchai, S.; Kojima, E.; Anderson, S.; Nomizu, M.; Tanaka, M.; Machida, M.; Date, T.; Toyota, K.; Ishida, S.; Watanabe, K.; et al.; *NP-06: a novel anti-human immunodeficiency virus polypeptide produced by a Streptomyces species*; *Antimicrob Agents Chemother* **1995**, *39*, 2345-2347.

[28] Gavrish, E.; Sit, C. S.; Cao, S.; Kandror, O.; Spoering, A.; Peoples, A.; Ling, L.; Fetterman, A.; Hughes, D.; Bissell, A.; Torrey, H.; Akopian, T.; Mueller, A.; Epstein, S.; Goldberg, A.; Clardy, J.; Lewis, K.; *Lassomycin, a ribosomally synthesized cyclic peptide, kills mycobacterium tuberculosis by targeting the ATP-dependent protease ClpC1P1P2*; *Chem Biol* **2014**, *21*, 509-518.

[29] Kersten, R. D.; Yang, Y. L.; Xu, Y.; Cimermancic, P.; Nam, S. J.; Fenical, W.; Fischbach, M. A.; Moore, B. S.; Dorrestein, P. C.; *A mass spectrometry-guided genome mining approach for natural product peptidogenomics*; *Nat Chem Biol* **2011**, *7*, 794-802.

[30] Maksimov, M. O.; Pelczer, I.; Link, A. J.; *Precursor-centric genome-mining approach for lasso peptide discovery*; *Proc Natl Acad Sci U S A* **2012**, *109*, 15223-15228.

[31] Wyss, D. F.; Lahm, H. W.; Manneberg, M.; Labhardt, A. M.; *Anantin--a peptide antagonist of the atrial natriuretic factor (ANF). II. Determination of the primary sequence by NMR on the basis of proton assignments*; *J Antibiot (Tokyo)* **1991**, *44*, 172-180.

[32] Salomon, R. A.; Farias, R. N.; *Microcin 25, a novel antimicrobial peptide produced by Escherichia coli*; *J Bacteriol* **1992**, *174*, 7428-7435.

[33] Yamasaki, M.; Yano, K.; Yoshida, M.; Matsuda, Y.; Yamaguchi, K.; *RES-701-1, a novel and*

selective endothelin type B receptor antagonist produced by Streptomyces sp. RE-701. II. Determination of the primary sequence; *J Antibiot (Tokyo)* **1994**, *47*, 276-280.

[34] Yano, K.; Toki, S.; Nakanishi, S.; Ochiai, K.; Ando, K.; Yoshida, M.; Matsuda, Y.; Yamasaki, M.; *MS-271, a novel inhibitor of calmodulin-activated myosin light chain kinase from Streptomyces sp.--I. Isolation, structural determination and biological properties of MS-271*; *Bioorg Med Chem* **1996**, *4*, 115-120.

[35] Blond, A.; Cheminant, M.; Destoumieux-Garzón, D.; Ségalas-Milazzo, I.; Peduzzi, J.; Goulard, C.; Rebuffat, S.; *Thermolysin-linearized microcin J25 retains the structured core of the native macrocyclic peptide and displays antimicrobial activity*; *European Journal of Biochemistry* **2002**, *269*, 6212-6222.

[36] Knappe, T. A.; Linne, U.; Robbel, L.; Marahiel, M. A.; *Insights into the biosynthesis and stability of the lasso peptide capistruin*; *Chem Biol* **2009**, *16*, 1290-1298.

[37] Hegemann, J. D.; Zimmermann, M.; Zhu, S.; Klug, D.; Marahiel, M. A.; *Lasso peptides from proteobacteria: Genome mining employing heterologous expression and mass spectrometry*; *Biopolymers* **2013**.

[38] Yano, K.; Yamasaki, M.; Yoshida, M.; Matsuda, Y.; Yamaguchi, K.; *RES-701-2, a novel and selective endothelin type B receptor antagonist produced by Streptomyces sp. II. Determination of the primary structure*; *J Antibiot (Tokyo)* **1995**, *48*, 1368-1370.

[39] Bellomio, A.; Vincent, P. A.; de Arcuri, B. F.; Salomon, R. A.; Morero, R. D.; Farias, R. N.; *The microcin J25 beta-hairpin region is important for antibiotic uptake but not for RNA polymerase and respiration inhibition*; *Biochem Biophys Res Commun* **2004**, *325*, 1454-1458.

[40] Helynck, G.; Dubertret, C.; Mayaux, J. F.; Leboul, J.; *Isolation of RP 71955, a new anti-HIV-1 peptide secondary metabolite*; *J Antibiot (Tokyo)* **1993**, *46*, 1756-1757.

[41] Kimura, K.; Kanou, F.; Takahashi, H.; Esumi, Y.; Uramoto, M.; Yoshihama, M.; *Propeptin, a new inhibitor of prolyl endopeptidase produced by Microbispora. I. Fermentation, isolation and biological properties*; *J Antibiot (Tokyo)* **1997**, *50*, 373-378.

[42] Esumi, Y.; Suzuki, Y.; Itoh, Y.; Uramoto, M.; Kimura, K.; Goto, M.; Yoshihama, M.; Ichikawa, T.; *Propeptin, a new inhibitor of prolyl endopeptidase produced by microbispora II. Determination of chemical structure*; *J Antibiot (Tokyo)* **2002**, *55*, 296-300.

[43] Tsunakawa, M.; Hu, S. L.; Hoshino, Y.; Detlefson, D. J.; Hill, S. E.; Furumai, T.; White, R. J.; Nishio, M.; Kawano, K.; Yamamoto, S.; et al.; *Siamycins I and II, new anti-HIV peptides: I. Fermentation, isolation, biological activity and initial characterization*; *J Antibiot (Tokyo)* **1995**, *48*, 433-434.

[44] Morishita, Y.; Chiba, S.; Tsukuda, E.; Tanaka, T.; Ogawa, T.; Yamasaki, M.; Yoshida, M.;

Kawamoto, I.; Matsuda, Y.; *RES-701-1, a novel and selective endothelin type B receptor antagonist produced by Streptomyces sp. RE-701. I. Characterization of producing strain, fermentation, isolation, physico-chemical and biological properties*; *J Antibiot (Tokyo)* **1994**, *47*, 269-275.

[45] Ogawa, T.; Ochiai, K.; Tanaka, T.; Tsukuda, E.; Chiba, S.; Yano, K.; Yamasaki, M.; Yoshida, M.; Matsuda, Y.; *RES-701-2, -3 and -4, novel and selective endothelin type B receptor antagonists produced by Streptomyces sp. I. Taxonomy of producing strains, fermentation, isolation, and biochemical properties*; *J Antibiot (Tokyo)* **1995**, *48*, 1213-1220.

[46] Weber, W.; Fischli, W.; Hochuli, E.; Kupfer, E.; Weibel, E. K.; *Anantin--a peptide antagonist of the atrial natriuretic factor (ANF). I. Producing organism, fermentation, isolation and biological activity*; *J Antibiot (Tokyo)* **1991**, *44*, 164-171.

[47] Zimmermann, M.; Hegemann, J. D.; Xie, X.; Marahiel, M. A.; *Characterization of Caulonodin Lasso Peptides Revealed Unprecedented N-Terminal Residues and a Precursor Motif Essential for Peptide Maturation*; *Chemical Science* **2014**.

[48] Potterat, O.; Wagner, K.; Gemmecker, G.; Mack, J.; Puder, C.; Vettermann, R.; Streicher, R.; *BI-32169, a bicyclic 19-peptide with strong glucagon receptor antagonist activity from Streptomyces sp*; *J Nat Prod* **2004**, *67*, 1528-1531.

[49] Rosengren, K. J.; Blond, A.; Afonso, C.; Tabet, J. C.; Rebuffat, S.; Craik, D. J.; *Structure of thermolysin cleaved microcin J25: extreme stability of a two-chain antimicrobial peptide devoid of covalent links*; *Biochemistry* **2004**, *43*, 4696-4702.

[50] Rebuffat, S.; Blond, A.; Destoumieux-Garzon, D.; Goulard, C.; Peduzzi, J.; *Microcin J25, from the macrocyclic to the lasso structure: implications for biosynthetic, evolutionary and biotechnological perspectives*; *Curr Protein Pept Sci* **2004**, *5*, 383-391.

[51] Blond, A.; Peduzzi, J.; Goulard, C.; Chiuchiolo, M. J.; Barthelemy, M.; Prigent, Y.; Salomon, R. A.; Farias, R. N.; Moreno, F.; Rebuffat, S.; *The cyclic structure of microcin J25, a 21-residue peptide antibiotic from Escherichia coli*; *Eur J Biochem* **1999**, *259*, 747-755.

[52] Blond, A.; Cheminant, M.; Segalas-Milazzo, I.; Peduzzi, J.; Barthelemy, M.; Goulard, C.; Salomon, R.; Moreno, F.; Farias, R.; Rebuffat, S.; *Solution structure of microcin J25, the single macrocyclic antimicrobial peptide from Escherichia coli*; *Eur J Biochem* **2001**, *268*, 2124-2133.

[53] Blond, A.; Cheminant, M.; Destoumieux-Garzon, D.; Segalas-Milazzo, I.; Peduzzi, J.; Goulard, C.; Rebuffat, S.; *Thermolysin-linearized microcin J25 retains the structured core of the native macrocyclic peptide and displays antimicrobial activity*; *Eur J Biochem* **2002**, *269*, 6212-6222.

[54] Severinov, K.; Semenova, E.; Kazakov, A.; Kazakov, T.; Gelfand, M. S.; *Low-molecular-weight post-translationally modified microcins*; *Mol Microbiol* **2007**, *65*, 1380-1394.

[55] Solbiati, J. O.; Ciaccio, M.; Farias, R. N.; Salomon, R. A.; *Genetic analysis of plasmid*

determinants for microcin J25 production and immunity; *J Bacteriol* **1996**, *178*, 3661-3663.

[56] Delgado, M. A.; Salomon, R. A.; *Molecular characterization of a DNA fragment carrying the basic replicon of pTUC100, the natural plasmid encoding the peptide antibiotic microcin J25 system*; *Plasmid* **2005**, *53*, 258-262.

[57] Inokoshi, J.; Matsuhama, M.; Miyake, M.; Ikeda, H.; Tomoda, H.; *Molecular cloning of the gene cluster for lariatin biosynthesis of Rhodococcus jostii K01-B0171*; *Appl Microbiol Biotechnol* **2012**, *95*, 451-460.

[58] Solbiati, J. O.; Ciaccio, M.; Farias, R. N.; Gonzalez-Pastor, J. E.; Moreno, F.; Salomon, R. A.; *Sequence analysis of the four plasmid genes required to produce the circular peptide antibiotic microcin J25*; *J Bacteriol* **1999**, *181*, 2659-2662.

[59] Yan, K. P.; Li, Y.; Zirah, S.; Goulard, C.; Knappe, T. A.; Marahiel, M. A.; Rebuffat, S.; *Dissecting the maturation steps of the lasso peptide microcin J25 in vitro*; *Chembiochem* **2012**, *13*, 1046-1052.

[60] Clarke, D. J.; Campopiano, D. J.; *Maturation of McjA precursor peptide into active microcin MccJ25*; *Org Biomol Chem* **2007**, *5*, 2564-2566.

[61] Pan, S. J.; Rajniak, J.; Cheung, W. L.; Link, A. J.; *Construction of a single polypeptide that matures and exports the lasso peptide microcin J25*; *Chembiochem* **2012**, *13*, 367-370.

[62] Pavlova, O.; Mukhopadhyay, J.; Sineva, E.; Ebright, R. H.; Severinov, K.; *Systematic structure-activity analysis of microcin J25*; *J Biol Chem* **2008**, *283*, 25589-25595.

[63] Cheung, W. L.; Pan, S. J.; Link, A. J.; *Much of the microcin J25 leader peptide is dispensable*; *J Am Chem Soc* **2010**, *132*, 2514-2515.

[64] Pan, S. J.; Rajniak, J.; Maksimov, M. O.; Link, A. J.; *The role of a conserved threonine residue in the leader peptide of lasso peptide precursors*; *Chem Commun (Camb)* **2012**, *48*, 1880-1882.

[65] Knappe, T. A.; Manzenrieder, F.; Mas-Moruno, C.; Linne, U.; Sasse, F.; Kessler, H.; Xie, X.; Marahiel, M. A.; *Introducing lasso peptides as molecular scaffolds for drug design: engineering of an integrin antagonist*; *Angew Chem Int Ed Engl* **2011**, *50*, 8714-8717.

[66] Pan, S. J.; Link, A. J.; *Sequence diversity in the lasso peptide framework: discovery of functional microcin J25 variants with multiple amino acid substitutions*; *J Am Chem Soc* **2011**, *133*, 5016-5023.

[67] Ducasse, R.; Yan, K. P.; Goulard, C.; Blond, A.; Li, Y.; Lescop, E.; Guittet, E.; Rebuffat, S.; Zirah, S.; *Sequence determinants governing the topology and biological activity of a lasso peptide, microcin J25*; *Chembiochem* **2012**, *13*, 371-380.

[68] Pomares, M. F.; Salomon, R. A.; Pavlova, O.; Severinov, K.; Farias, R.; Vincent, P. A.; *Potential applicability of chymotrypsin-susceptible microcin J25 derivatives to food preservation*;

Appl Environ Microbiol **2009**, *75*, 5734-5738.

[69] Hegemann, J. D.; De Simone, M.; Zimmermann, M.; Knappe, T. A.; Xie, X.; Di Leva, F. S.; Marinelli, L.; Novellino, E.; Zahler, S.; Kessler, H.; Marahiel, M. A.; *Rational Improvement of the Affinity and Selectivity of Integrin Binding of Grafted Lasso Peptides*; *Journal of Medicinal Chemistry* **2014**, *57*, 5829-5834.

[70] Chiuchiolo, M. J.; Delgado, M. A.; Farias, R. N.; Salomon, R. A.; *Growth-phase-dependent expression of the cyclopeptide antibiotic microcin J25*; *J Bacteriol* **2001**, *183*, 1755-1764.

[71] Semenova, E.; Yuzenkova, Y.; Peduzzi, J.; Rebuffat, S.; Severinov, K.; *Structure-activity analysis of microcinJ25: distinct parts of the threaded lasso molecule are responsible for interaction with bacterial RNA polymerase*; *J Bacteriol* **2005**, *187*, 3859-3863.

[72] Yuzenkova, J.; Delgado, M.; Nechaev, S.; Savalia, D.; Epshtein, V.; Artsimovitch, I.; Mooney, R. A.; Landick, R.; Farias, R. N.; Salomon, R.; Severinov, K.; *Mutations of bacterial RNA polymerase leading to resistance to microcin j25*; *J Biol Chem* **2002**, *277*, 50867-50875.

[73] Adelman, K.; Yuzenkova, J.; La Porta, A.; Zenkin, N.; Lee, J.; Lis, J. T.; Borukhov, S.; Wang, M. D.; Severinov, K.; *Molecular mechanism of transcription inhibition by peptide antibiotic Microcin J25*; *Mol Cell* **2004**, *14*, 753-762.

[74] Kuznedelov, K.; Semenova, E.; Knappe, T. A.; Mukhamedyarov, D.; Srivastava, A.; Chatterjee, S.; Ebright, R. H.; Marahiel, M. A.; Severinov, K.; *The antibacterial threaded-lasso peptide capistruin inhibits bacterial RNA polymerase*; *J Mol Biol* **2011**, *412*, 842-848.

[75] Delgado, M. A.; Rintoul, M. R.; Farias, R. N.; Salomon, R. A.; *Escherichia coli RNA polymerase is the target of the cyclopeptide antibiotic microcin J25*; *J Bacteriol* **2001**, *183*, 4543-4550.

[76] Salomon, R. A.; Farias, R. N.; *The FhuA protein is involved in microcin 25 uptake*; *J Bacteriol* **1993**, *175*, 7741-7742.

[77] Salomon, R. A.; Farias, R. N.; *The peptide antibiotic microcin 25 is imported through the TonB pathway and the SbmA protein*; *J Bacteriol* **1995**, *177*, 3323-3325.

[78] Mathavan, I.; Zirah, S.; Mehmood, S.; Choudhury, H. G.; Goulard, C.; Li, Y.; Robinson, C. V.; Rebuffat, S.; Beis, K.; *Structural basis for hijacking siderophore receptors by antimicrobial lasso peptides*; *Nat Chem Biol* **2014**, *10*, 340-342.

[79] Rintoul, M. R.; de Arcuri, B. F.; Salomon, R. A.; Farias, R. N.; Morero, R. D.; *The antibacterial action of microcin J25: evidence for disruption of cytoplasmic membrane energization in Salmonella newport*; *FEMS Microbiol Lett* **2001**, *204*, 265-270.

[80] Vincent, P. A.; Delgado, M. A.; Farias, R. N.; Salomon, R. A.; *Inhibition of Salmonella enterica serovars by microcin J25*; *FEMS Microbiol Lett* **2004**, *236*, 103-107.

[81] Bellomio, A.; Vincent, P. A.; de Arcuri, B. F.; Farias, R. N.; Morero, R. D.; *Microcin J25 has dual and independent mechanisms of action in Escherichia coli: RNA polymerase inhibition and increased superoxide production*; *J Bacteriol* **2007**, *189*, 4180-4186.

[82] Vincent, P. A.; Bellomio, A.; de Arcuri, B. F.; Farias, R. N.; Morero, R. D.; *MccJ25 C-terminal is involved in RNA-polymerase inhibition but not in respiration inhibition*; *Biochem Biophys Res Commun* **2005**, *331*, 549-551.

[83] Chalon, M. C.; Bellomio, A.; Solbiati, J. O.; Morero, R. D.; Farias, R. N.; Vincent, P. A.; *Tyrosine 9 is the key amino acid in microcin J25 superoxide overproduction*; *FEMS Microbiol Lett* **2009**, *300*, 90-96.

[84] Braun, V.; Mahren, S.; Sauter, A.; *Gene regulation by transmembrane signaling*; *Biometals* **2006**, *19*, 103-113.

[85] Breidenstein, E.; Mahren, S.; Braun, V.; *Residues involved in FecR binding are localized on one side of the FecA signaling domain in Escherichia coli*; *J Bacteriol* **2006**, *188*, 6440-6442.

[86] Braun, V.; Herrmann, C.; *Docking of the periplasmic FecB binding protein to the FecCD transmembrane proteins in the ferric citrate transport system of Escherichia coli*; *J Bacteriol* **2007**, *189*, 6913-6918.

[87] Weckbecker, G.; Lewis, I.; Albert, R.; Schmid, H. A.; Hoyer, D.; Bruns, C.; *Opportunities in somatostatin research: biological, chemical and therapeutic aspects*; *Nat Rev Drug Discov* **2003**, *2*, 999-1017.

[88] Mankoff, D. A.; Link, J. M.; Linden, H. M.; Sundararajan, L.; Krohn, K. A.; *Tumor receptor imaging*; *J Nucl Med* **2008**, *49 Suppl 2*, 149S-163S.

[89] Oberg, K. E.; Reubi, J. C.; Kwekkeboom, D. J.; Krenning, E. P.; *Role of somatostatins in gastroenteropancreatic neuroendocrine tumor development and therapy*; *Gastroenterology* **2010**, *139*, 742-753, 753 e741.

[90] Zoller, F.; Haberkorn, U.; Mier, W.; *Miniproteins as phage display-scaffolds for clinical applications*; *Molecules* **2011**, *16*, 2467-2485.

[91] Fani, M.; Maecke, H. R.; Okarvi, S. M.; *Radiolabeled peptides: valuable tools for the detection and treatment of cancer*; *Theranostics* **2012**, *2*, 481-501.

[92] Barbieri, F.; Bajetto, A.; Pattarozzi, A.; Gatti, M.; Wurth, R.; Thellung, S.; Corsaro, A.; Villa, V.; Nizzari, M.; Florio, T.; *Peptide receptor targeting in cancer: the somatostatin paradigm*; *Int J Pept* **2013**, *2013*, 926295.

[93] Koehbach, J.; O'Brien, M.; Muttenthaler, M.; Miazzo, M.; Akcan, M.; Elliott, A. G.; Daly, N. L.; Harvey, P. J.; Arrowsmith, S.; Gunasekera, S.; Smith, T. J.; Wray, S.; Goransson, U.; Dawson, P. E.; Craik, D. J.; Freissmuth, M.; Gruber, C. W.; *Oxytocic plant cyclotides as templates for peptide*

G protein-coupled receptor ligand design; *Proc Natl Acad Sci U S A* **2013**, *110*, 21183-21188.

[94] Schulz, S.; Lehmann, A.; Kliewer, A.; Nagel, F.; *Fine-tuning Somatostatin Receptor Signaling by Agonist-Selective Phosphorylation and Dephosphorylation*; *Br J Pharmacol* **2013**.

[95] Wang, C. K.; Gruber, C. W.; Cemazar, M.; Siatskas, C.; Tagore, P.; Payne, N.; Sun, G.; Wang, S.; Bernard, C. C.; Craik, D. J.; *Molecular Grafting onto a Stable Framework Yields Novel Cyclic Peptides for the Treatment of Multiple Sclerosis*; *ACS Chemical Biology* **2013**, *9*, 156-163.

[96] Marugan, J. J.; Manthey, C.; Anaclerio, B.; Lafrance, L.; Lu, T.; Markotan, T.; Leonard, K. A.; Crysler, C.; Eisennagel, S.; Dasgupta, M.; Tomczuk, B.; *Design, synthesis, and biological evaluation of novel potent and selective alphavbeta3/alphavbeta5 integrin dual inhibitors with improved bioavailability. Selection of the molecular core*; *J Med Chem* **2005**, *48*, 926-934.

[97] Sato, A. K.; Viswanathan, M.; Kent, R. B.; Wood, C. R.; *Therapeutic peptides: technological advances driving peptides into development*; *Curr Opin Biotechnol* **2006**, *17*, 638-642.

[98] Stragies, R.; Osterkamp, F.; Zischinsky, G.; Vossmeyer, D.; Kalkhof, H.; Reimer, U.; Zahn, G.; *Design and synthesis of a new class of selective integrin alpha5beta1 antagonists*; *J Med Chem* **2007**, *50*, 3786-3794.

[99] Avraamides, C. J.; Garmy-Susini, B.; Varner, J. A.; *Integrins in angiogenesis and lymphangiogenesis*; *Nat Rev Cancer* **2008**, *8*, 604-617.

[100] Biron, E.; Chatterjee, J.; Ovadia, O.; Langenegger, D.; Brueggen, J.; Hoyer, D.; Schmid, H. A.; Jelinek, R.; Gilon, C.; Hoffman, A.; Kessler, H.; *Improving oral bioavailability of peptides by multiple N-methylation: somatostatin analogues*; *Angew Chem Int Ed Engl* **2008**, *47*, 2595-2599.

[101] Antosova, Z.; Mackova, M.; Kral, V.; Macek, T.; *Therapeutic application of peptides and proteins: parenteral forever?*; *Trends Biotechnol* **2009**, *27*, 628-635.

[102] Pazgier, M.; Liu, M.; Zou, G.; Yuan, W.; Li, C.; Li, C.; Li, J.; Monbo, J.; Zella, D.; Tarasov, S. G.; Lu, W.; *Structural basis for high-affinity peptide inhibition of p53 interactions with MDM2 and MDMX*; *Proc Natl Acad Sci U S A* **2009**, *106*, 4665-4670.

[103] Desgrosellier, J. S.; Cheresh, D. A.; *Integrins in cancer: biological implications and therapeutic opportunities*; *Nat Rev Cancer* **2010**, *10*, 9-22.

[104] Mas-Moruno, C.; Rechenmacher, F.; Kessler, H.; *Cilengitide: the first anti-angiogenic small molecule drug candidate design, synthesis and clinical evaluation*; *Anticancer Agents Med Chem* **2010**, *10*, 753-768.

[105] Chan, L. Y.; Gunasekera, S.; Henriques, S. T.; Worth, N. F.; Le, S. J.; Clark, R. J.; Campbell, J. H.; Craik, D. J.; Daly, N. L.; *Engineering pro-angiogenic peptides using stable, disulfide-rich cyclic scaffolds*; *Blood* **2011**, *118*, 6709-6717.

[106] Young, T. S.; Young, D. D.; Ahmad, I.; Louis, J. M.; Benkovic, S. J.; Schultz, P. G.; *Evolution*

of cyclic peptide protease inhibitors; *Proc Natl Acad Sci U S A* **2011**, *108*, 11052-11056.

[107] Goodman, S. L.; Picard, M.; *Integrins as therapeutic targets*; *Trends Pharmacol Sci* **2012**, *33*, 405-412.

[108] Ji, Y.; Majumder, S.; Millard, M.; Borra, R.; Bi, T.; Elnagar, A. Y.; Neamati, N.; Shekhtman, A.; Camarero, J. A.; *In vivo activation of the p53 tumor suppressor pathway by an engineered cyclotide*; *J Am Chem Soc* **2013**, *135*, 11623-11633.

[109] Kapp, T. G.; Rechenmacher, F.; Sobahi, T. R.; Kessler, H.; *Integrin modulators: a patent review*; *Expert Opin Ther Pat* **2013**, *23*, 1273-1295.

[110] Poth, A. G.; Chan, L. Y.; Craik, D. J.; *Cyclotides as grafting frameworks for protein engineering and drug design applications*; *Biopolymers* **2013**, *100*, 480-491.

[111] Conibear, A. C.; Bochen, A.; Rosengren, K. J.; Stupar, P.; Wang, C.; Kessler, H.; Craik, D. J.; *The Cyclic Cystine Ladder of Theta-Defensins as a Stable, Bifunctional Scaffold: A Proof-of-Concept Study Using the Integrin-Binding RGD Motif*; *Chembiochem* **2014**, *15*, 451-459.

[112] Dechantsreiter, M. A.; Planker, E.; Matha, B.; Lohof, E.; Holzemann, G.; Jonczyk, A.; Goodman, S. L.; Kessler, H.; *N-Methylated cyclic RGD peptides as highly active and selective alpha(V)beta(3) integrin antagonists*; *J Med Chem* **1999**, *42*, 3033-3040.

[113] Patel, Y. C.; Wheatley, T.; *In vivo and in vitro plasma disappearance and metabolism of somatostatin-28 and somatostatin-14 in the rat*; *Endocrinology* **1983**, *112*, 220-225.

[114] Aumailley, M.; Gurrath, M.; Muller, G.; Calvete, J.; Timpl, R.; Kessler, H.; *Arg-Gly-Asp constrained within cyclic pentapeptides. Strong and selective inhibitors of cell adhesion to vitronectin and laminin fragment P1*; *FEBS Lett* **1991**, *291*, 50-54.

[115] Gurrath, M.; Muller, G.; Kessler, H.; Aumailley, M.; Timpl, R.; *Conformation/activity studies of rationally designed potent anti-adhesive RGD peptides*; *Eur J Biochem* **1992**, *210*, 911-921.

[116] Haubner, R.; Gratias, R.; Diefenbach, B.; Goodman, S. L.; Jonczyk, A.; Kessler, H.; *Structural and Functional Aspects of RGD-Containing Cyclic Pentapeptides as Highly Potent and Selective Integrin αVβ3 Antagonists*; *Journal of the American Chemical Society* **1996**, *118*, 7461-7472.

[117] Rosengren, K. J.; Daly, N. L.; Plan, M. R.; Waine, C.; Craik, D. J.; *Twists, knots, and rings in proteins. Structural definition of the cyclotide framework*; *J Biol Chem* **2003**, *278*, 8606-8616.

[118] Gunasekera, S.; Foley, F. M.; Clark, R. J.; Sando, L.; Fabri, L. J.; Craik, D. J.; Daly, N. L.; *Engineering stabilized vascular endothelial growth factor-A antagonists: synthesis, structural characterization, and bioactivity of grafted analogues of cyclotides*; *J Med Chem* **2008**, *51*, 7697-7704.

[119] Thongyoo, P.; Bonomelli, C.; Leatherbarrow, R. J.; Tate, E. W.; *Potent inhibitors of beta-*

tryptase and human leukocyte elastase based on the MCoTI-II scaffold; *J Med Chem* **2009**, *52*, 6197-6200.

[120] Sommerhoff, C. P.; Avrutina, O.; Schmoldt, H. U.; Gabrijelcic-Geiger, D.; Diederichsen, U.; Kolmar, H.; *Engineered cystine knot miniproteins as potent inhibitors of human mast cell tryptase beta*; *J Mol Biol* **2010**, *395*, 167-175.

[121] Eliasen, R.; Daly, N. L.; Wulff, B. S.; Andresen, T. L.; Conde-Frieboes, K. W.; Craik, D. J.; *Design, synthesis, structural and functional characterization of novel melanocortin agonists based on the cyclotide kalata B1*; *J Biol Chem* **2012**, *287*, 40493-40501.

[122] Kimura, R. H.; Teed, R.; Hackel, B. J.; Pysz, M. A.; Chuang, C. Z.; Sathirachinda, A.; Willmann, J. K.; Gambhir, S. S.; *Pharmacokinetically stabilized cystine knot peptides that bind alpha-v-beta-6 integrin with single-digit nanomolar affinities for detection of pancreatic cancer*; *Clin Cancer Res* **2012**, *18*, 839-849.

[123] Wong, C. T.; Rowlands, D. K.; Wong, C. H.; Lo, T. W.; Nguyen, G. K.; Li, H. Y.; Tam, J. P.; *Orally active peptidic bradykinin B1 receptor antagonists engineered from a cyclotide scaffold for inflammatory pain treatment*; *Angew Chem Int Ed Engl* **2012**, *51*, 5620-5624.

[124] Jagadish, K.; Borra, R.; Lacey, V.; Majumder, S.; Shekhtman, A.; Wang, L.; Camarero, J. A.; *Expression of fluorescent cyclotides using protein trans-splicing for easy monitoring of cyclotide-protein interactions*; *Angew Chem Int Ed Engl* **2013**, *52*, 3126-3131.

[125] Pan, S. J.; Cheung, W. L.; Link, A. J.; *Engineered gene clusters for the production of the antimicrobial peptide microcin J25*; *Protein Expr Purif* **2010**, *71*, 200-206.

[126] Kimura, K.; Yamazaki, M.; Sasaki, N.; Yamashita, T.; Negishi, S.; Nakamura, T.; Koshino, H.; *Novel propeptin analog, propeptin-2, missing two amino acid residues from the propeptin C-terminus loses antibiotic potency*; *J Antibiot (Tokyo)* **2007**, *60*, 519-523.

[127] Gibson, H. W.; Nagvekar, D. S.; Yamaguchi, N.; Bhattacharjee, S.; Wang, H.; Vergne, M. J.; Hercules, D. M.; *Polyamide Pseudorotaxanes, Rotaxanes, and Catenanes Based on Bis(5-carboxy-1,3-phenylene)-(3x+2)-crown-x Ethers*; *Macromolecules* **2004**, *37*, 7514-7529.

[128] Dasgupta, S.; Huang, K. W.; Wu, J.; *Trifluoromethyl acting as stopper in [2]rotaxane*; *Chem Commun (Camb)* **2012**, *48*, 4821-4823.

[129] Dasgupta, S.; Wu, J.; *Formation of [2]rotaxanes by encircling [20], [21] and [22]crown ethers onto the dibenzylammonium dumbbell*; *Chemical Science* **2012**, *3*, 425-432.

[130] Xiong, J. P.; Stehle, T.; Zhang, R.; Joachimiak, A.; Frech, M.; Goodman, S. L.; Arnaout, M. A.; *Crystal structure of the extracellular segment of integrin alpha Vbeta3 in complex with an Arg-Gly-Asp ligand*; *Science* **2002**, *296*, 151-155.

[131] Kim, L.; Mogk, A.; Schumann, W.; *A xylose-inducible Bacillus subtilis integration vector and*

its application; *Gene* **1996**, *181*, 71-76.
[132] Gomez-Escribano, J. P.; Bibb, M. J.; *Engineering Streptomyces coelicolor for heterologous expression of secondary metabolite gene clusters*; *Microb Biotechnol* **2011**, *4*, 207-215.
[133] Kuipers, O. P.; de Ruyter, P. G. G. A.; Kleerebezem, M.; de Vos, W. M.; *Quorum sensing-controlled gene expression in lactic acid bacteria*; *Journal of Biotechnology* **1998**, *64*, 15-21.

7. Acknowledgment

First, I would like to thank Prof. Dr. Mohamed A. Marahiel for not only giving me the opportunity to pursue my PhD in his group, but also for his continuous support and encouragement throughout my work. Furthermore, I would like to thank him for the trust he put into me and my ideas and for helping me to become the scientist I am today. I am especially grateful for the multitude of opportunities where he enabled me to not only attend to intriguing and inspiring scientific meetings, but also to present my research there to the scientific community. I also greatly appreciate that he always took the time to discuss problems of any kind with me and that he treated me respectfully, even when we disagreed on a subject matter.

I would also like to extend my thanks to Prof. Dr. Peter Graumann for not only being a part of my thesis committee, but also for his willingness to review this thesis.

I would additionally like to thank the other members of my thesis committee, Prof. Dr. Wolfgang Buckel and Prof. Dr. Norbert Hampp, for their interest in my work and their time.

Another person I would like to thank is Dr. Xiulan Xie and the great effort and enthusiasm with which she conducted the NMR based structure elucidations of our lasso peptides. Her work is not only an essential corner stone for all our publications, but the friendly and fruitful discussions with her were always greatly appreciated.

Furthermore, I would like to express my gratitude towards Dr. Uwe Linne and the other members of the MS facilities. Without them, it would not have been possible to gather and evaluate all the MS data presented in all our studies and without the constant technical assistance of Uwe, all the problems with our HPLC and LC-MS systems would have driven me insane by now.

I would also like to thank Ralf Pöschke, Dr. Holger Steuber and Dr. Klaus Harms for their assistance during the crystallization and X-ray structure elucidation of our lasso peptides.

Additionally, I would like to extend my gratitude towards Prof. Dr. Sylvie Rebuffat, Dr. Yanyan Li, Dr. Séverine Zirah and all the other participants of our annual lasso peptide meetings. Their input and the meetings in general were always a great source of inspiration and independent if we met in Paris or Marburg, these meetings are always associated with lots of fond memories.

I also gratefully acknowledge the help and assistance of Dr. Mariarosaria De Simone, Francesco Saverio Di Leva, Dr. Luciana Marinelli, Prof. Dr. Ettore Novellino, Prof. Dr. Stefan Zahler and Prof. Dr. Horst Kessler. All of them were crucial for the success of our epitope grafting project and it was a pleasure to work together with such a large group of interested and talented scientists.

Moreover, I would like to thank Antje Schäfer and Kristin Ried for their help with some of my projects and their technical assistance in general. The additional technical assistance of Christiane Bomm and Gabriele Schimpff-Weiland is also gratefully acknowledged.

Many thanks also go to my former lab internship students Lara Golde, Elena Schütz, Mohamed Bouygheoussan, Sabrina Henche, Sabine Düwel and Julian Koch. Their work was not only a great help during my time as PhD student, but it was also a great joy for me to teach and cooperate with each of them.

My deepest thanks go to all the current and former members of the Marahiel group and everyone who in some way or the other joined the AK66 during my time here. Not only was the time spent and all the discussions I had here a rich source of ideas and inspiration, but the consistently nice and welcoming atmosphere in this group made the lab more than just a workplace.

Finally, I would like to extend my most heartfelt gratitude towards all of my close friends and my family. Your constant support through all the good times we shared as well as through all the periods of tedious and late work hours was a major driving force that kept me going. For this and other reasons, I would like to dedicate this thesis to all of you.

8. Appendix

This appendix contains the Supporting Information of all original publications that were attributed to this thesis. These are namely

Julian D. Hegemann, Marcel Zimmermann, Xiulan Xie, and Mohamed A. Marahiel, *Caulosegnins I-III: A Highly Diverse Group of Lasso Peptides Derived From a Single Biosynthetic Gene Cluster*, *J. Am. Chem. Soc.* **2013**, *135(1)*, 210-222.

Marcel Zimmermann, Julian D. Hegemann, Xiulan Xie, and Mohamed A. Marahiel, *The Astexin-1 Lasso Peptides: Biosynthesis, Stability, and Structural Studies*, *Chem. Biol.* **2013**, *20(4)*, 558-569.

Julian D. Hegemann, Marcel Zimmermann, Shaozhou Zhu, Dennis Klug, Mohamed A. Marahiel, *Lasso Peptides From Proteobacteria: Genome Mining Employing Heterologous Expression and Mass Spectrometry*, *Biopolymers Pept. Sci.* **2013**, *100(5)*, 527-542.

Julian D. Hegemann, Marcel Zimmermann, Shaozhou Zhu, Holger Steuber, Klaus Harms, Xiulan Xie, Mohamed A. Marahiel, Xanthomonins I-III: A New Class of Lasso Peptides With a Seven-Residue Macrolactam Ring, *Angew. Chem. Int. Ed. Engl.* **2014**, *53(8)*, 2230-2234 and *Angew. Chem.* **2014**, 126(8), 2262-2266.

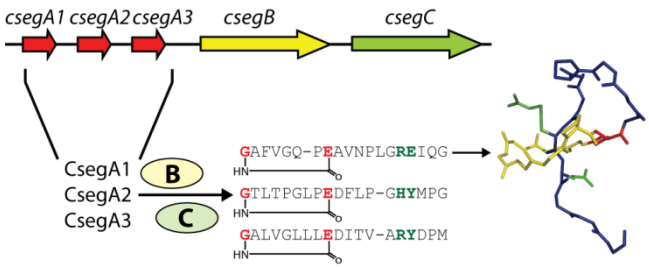
Marcel Zimmermann, Julian D. Hegemann, Xiulan Xie, and Mohamed A. Marahiel, *Characterization of Caulonodin Lasso Peptides Revealed Unprecedented N-Terminal Residues and a Precursor Motif Essential for Peptide Maturation*, *Chem. Sci.* **2014**, *Epub ahead of print*.
(doi: 10.1039/c4sc01428f)

Julian D. Hegemann, Mariarosaria De Simone, Marcel Zimmermann, Thomas A. Knappe, Xiulan Xie, Francesco Saverio Di Leva, Luciana Marinelli, Ettore Novellino, Stefan Zahler, Horst Kessler, and Mohamed A. Marahiel, *Rational Improvement of the Affinity and Selectivity of Integrin Binding of Grafted Lasso Peptides*, *J. Med. Chem.* **2014**, *57(13)*, 5829-5834.

Caulosegnins I-III: A Highly Diverse Group of Lasso Peptides Derived from a Single Biosynthetic Gene Cluster

*Julian D. Hegemann, Marcel Zimmerman, Xiulan Xie and Mohamed A. Marahiel**

Department of Chemistry, Philipps-University Marburg, Hans-Meerwein-Strasse, D-35032, Marburg, Germany



Supporting Information

Protease Stability Assays Highlight the Resistance of the Lasso Fold Against Proteolytic Degradation. The stability of the caulosegnins against proteolytic degradation by trypsin, chymotrypsin, elastase, proteinase K and carboxypeptidase Y was tested. These assays were also performed with samples of caulosegnin I and III which were mostly unthreaded by heating them to 95 °C for 1 h. This was done to allow a comparison of the lasso fold to the branched cyclic analogues in regard to their protease sensitivity and to further proof their proposed topologies.

In general these experiments showed the high stability of caulosegnins against proteolytic degradation with caulosegnin I and III being only slightly digested in some cases, while caulosegnin II was not affected at all by any of the tested proteases. On the other hand, the unthreaded caulosegin I and III were highly sensitive against all tested proteases (Figure S11-S13, Table S8).

As carboxypeptidase Y cleaves off amino acids from the C-terminus of a peptide chain, these assays were of particular interest as they could provide information about the present topology of the digested compound. The branched cyclic analogue of caulosegnin I was readily digested to a point where only four amino acids after the macrolactam ring remained, while the branched cyclic analogue of caulosegnin III was digested until only one additional amino acid after the macrolactam ring was present, thus proving the proposed topology in both cases. In contrast the assays of caulosegnin I and III showed that even after 16 h of incubation only trace amounts of these degradation products were present, which were most likely formed by the digestion of branched cyclic peptide formed during the course of these assays because of the heat lability of both lasso peptides. These observations are in accordance with the notion that the macrolactam ring is able to sterically shield amino acids positioned inside or close by the ring from proteolytic degradation.

Mass Spectrometric Analysis Supports the Suggested Plugs of Caulosegnin II and III. MS^2 analysis was used to confirm the primary structures of the investigated lasso peptides, their variants and their branched cyclic analogues. MS^2 spectra, independent from the present topology, contained all fragments corresponding to the predicted amino acid sequence (lacking a water molecule) up to the position of the proposed macrolactam ring. This is not only a strong hint for the presence of a macrolactam ring, but also indicates that it has to be formed between the suggested residues. Tandem mass spectrometry thus confirms the predicted primary structure, indicates that the proposed position and presence of a macrolactam ring is correct and shows a difference between the proposed lasso and branched cyclic topologies. This is, in combination with the facts that the predictions were all based on information derived from the biosynthetic gene cluster and that NMR structure elucidation confirmed the lasso topology of caulosegnin I, strong evidence of all caulosegnins being true lasso peptides. With a similar set of MS data, but lacking a biochemical link between the compound and the proposed biosynthetic gene cluster, any mutational analysis, protease assays, thermal stability assays and without any 3D structural proof of the biosynthetic gene cluster being generally able to produce lasso peptides, the discovery of two new lasso peptides was recently reported (Kersten, R. D.; Yang, Y. L.; Xu, Y.; Cimermancic, P.; Nam, S. J.; Fenical, W.; Fischbach, M. A.; Moore, B. S.; Dorrestein, P. C. *Nat. Chem. Biol.* **2011**, *7,* 794-802), which further emphasizes the potency of mass spectrometric analysis in this regard.

Additionally comparison of signal intensities of a lasso peptide and its branched cycled analogue yields interesting information about the lasso fold. In case of caulosegnin I (Figure S15) relative signal intensities for both the single and double charged b17 fragment ions increased drastically for the branched cyclic compared to the lasso peptide. This fragment corresponds to a bond breakage after Ile17, which is located one amino acid after the lower plug Glu16. In a similar fashion the intensities of the b17 fragment ions of caulosegnin II (Figure S16) and caulosegnin III (Figure S17) show a significant increase for the branched cylic analogues, which corresponds to bond breakages in the same distance to the suggested plugs and thus supports the idea of Tyr16 sterically trapping the

tail inside the ring of the lasso peptides.

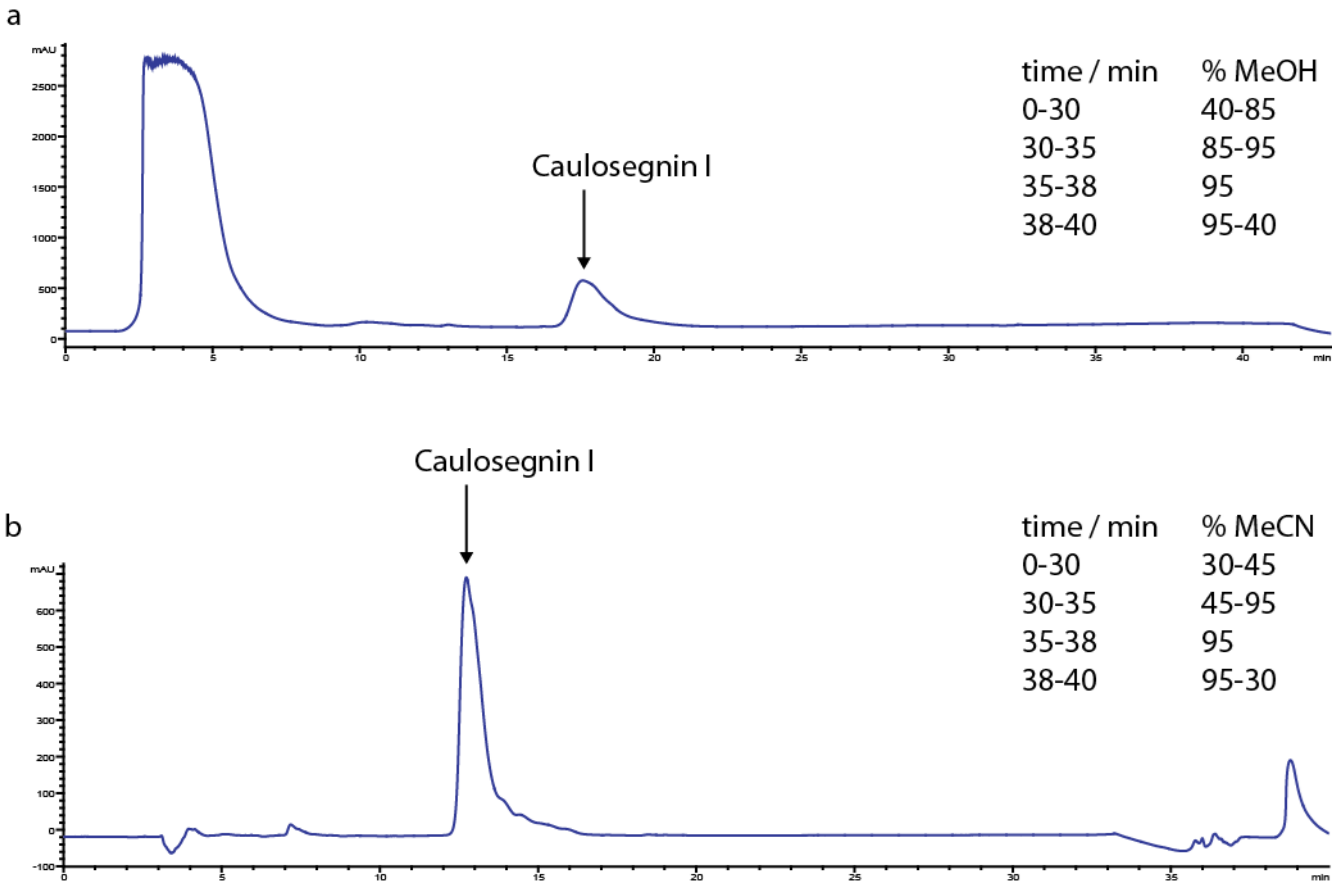
For caulosegnin II it can additionally be observed, that the lasso peptide fragmentation yields much more intensive signals for the fragments formed by bond breakage before the Pro13 (namely the y7 and b12 fragment ions). This suggests a constrained conformation of Pro13 in the lasso fold, leading to the notion of Pro13 possibly being the turning point of the loop of caulosegnin II similar to Pro12 being the turning point of the loop of caulosegnin I. Not surprisingly a similar observation is not found for caulosegnin I, as the additional amino acid in the loop would lead to a more relaxed conformation.

In the case of caulosegnin III bond breakages before the Ala14 (leading to the y6 and b13 fragment ions) are also more intensive for the lasso peptide, which suggests that because of the lacking proline residue, a slightly different, but still constrained, loop conformation is present. Furthermore the fragmentation spectrum of the branched cyclic analogue of caulosegnin III exhibits much stronger signals for the b14, b15, b16 and y5 fragment ions. These are all ions formed by a bond breakage in the area where the tail supposedly passes the ring in the lasso peptide. As it can be expected that bond breakages occur more often if the peptide bond is not shielded by the macrolactam ring, these are additional hints for the caulosegnin III plug being in the assumed area of the peptide's tail.

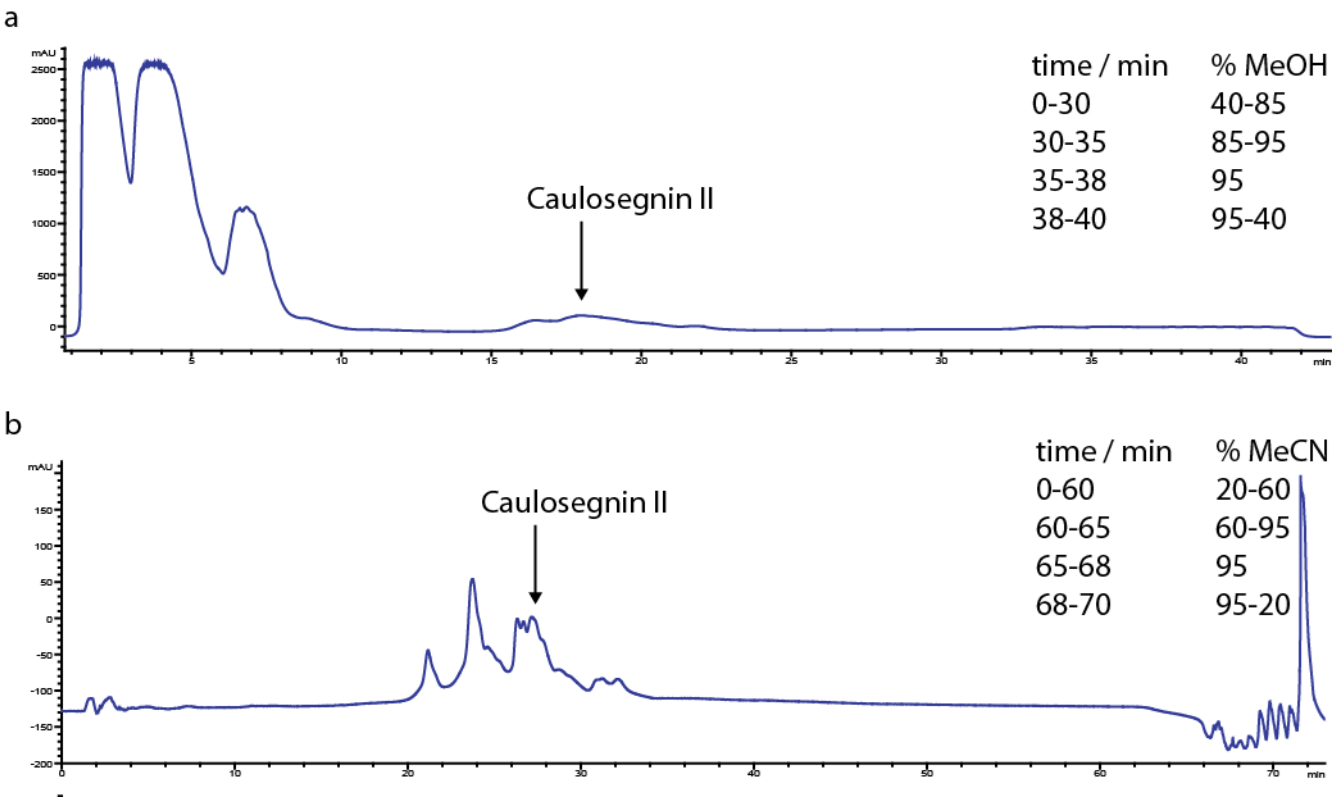
NMR Spectroscopic Analysis: Torsion Angles and Side Chain Stereospecific Assignment. Distinct values of vicinal coupling constants (6.0 and 9.7 Hz) were observed for Gly1, Gly5, and Gly14, thus the torsion angles ϕ were unequivocally assigned to -120 ± 15° (Güntert, P.; Braun, W.; Billeter, M.; Wüthrich, K. *J. Am. Chem. Soc.* **1989**, *111*, 3997-4004). The torsion angles ϕ were restrained to -70° ± 30° for Gln6, Glu8, Ala9, Asn11, Leu13, Glu16, Ile17, and Gln18 with $^{3}J_{HN\alpha} < 9$ Hz, and -120° ± 30° for the remaining residues with $^{3}J_{HN\alpha} \geq 9$ Hz. Stereospecific assignment of the following prochiral β–methylene protons were fulfilled by measuring $^{3}J\alpha\beta$ and analyzing patterns of the intraresidual NOE interactions $d_{\alpha\beta}$ and $d_{N\beta}$ (Wagner, G. *Prog. NMR Spec*. **1990**, *22*, 101-139.): Phe3, Glu8, Leu13, Glu16, and Gln18 in (t^2g^3); Gln6 and Arg15 in (g^2g^3), and Asn11 in (g^2t^3). For

t^2g^3, g^2g^3, and g^2t^3 conformations around the Cα–Cβ bond the torsion angle χ^1 was constrained in the range of -60 ± 30°, 60 ± 30°, and 150 ± 30°, respectively.

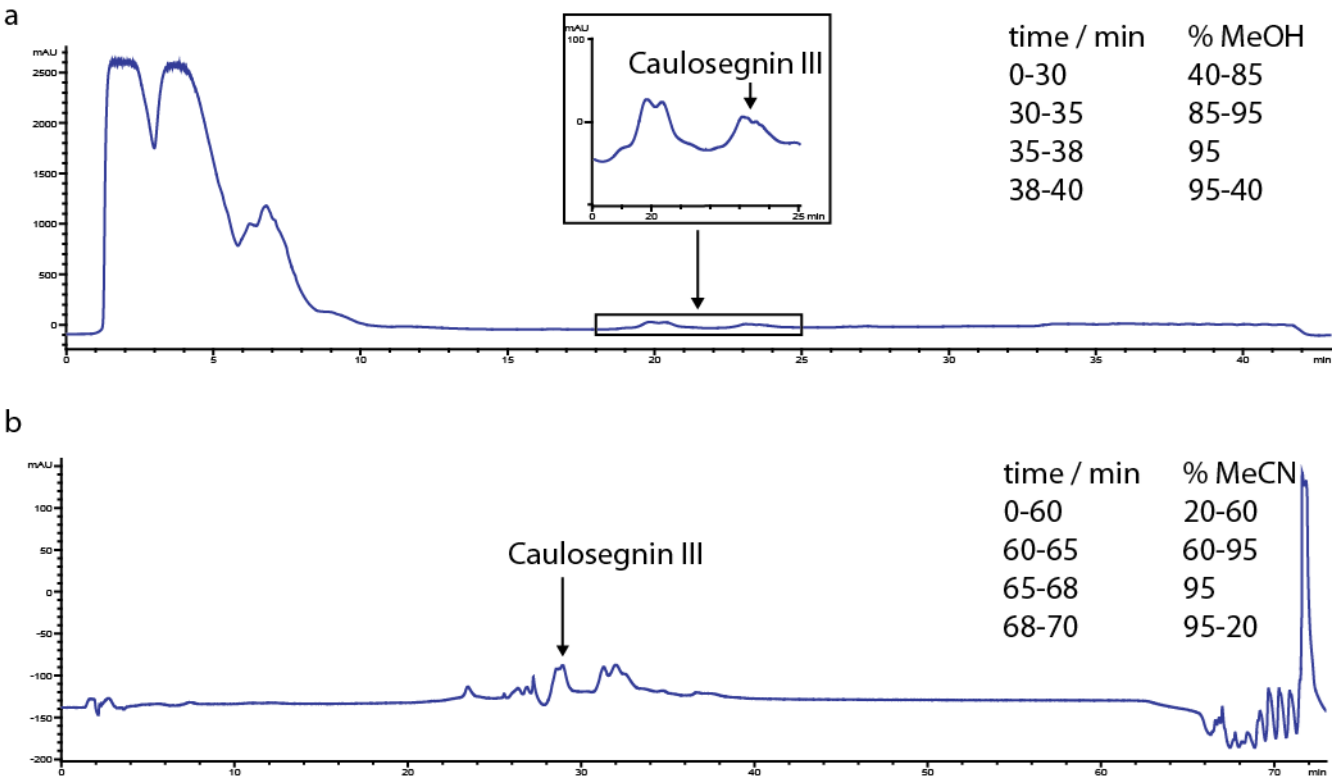
Supplementary Figures and Tables



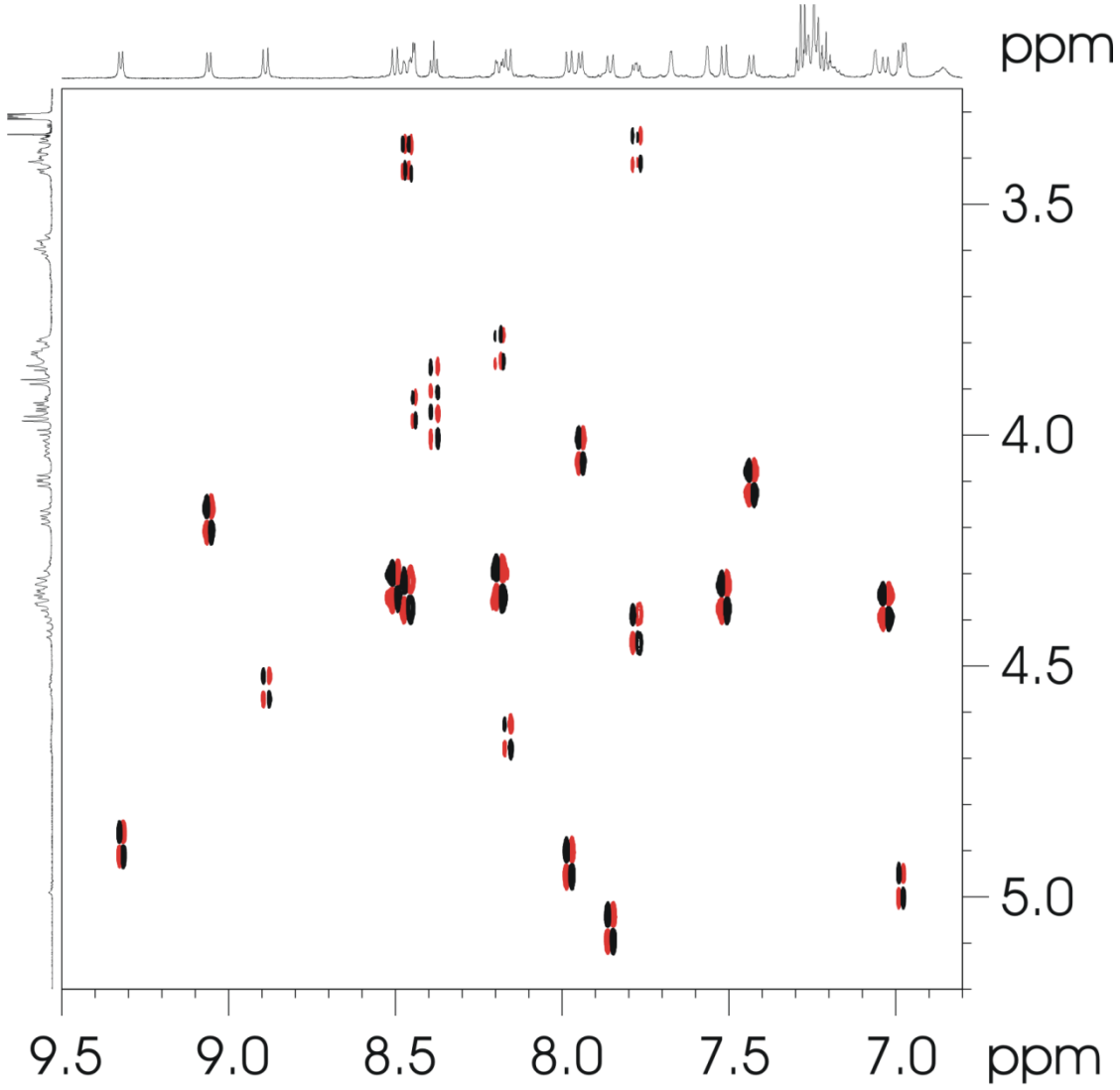
Supplementary Figure S1. Chromatograms of the HPLC purification of caulosegnin I. (a) First purification with water/0.05% formic acid and MeOH/0.045% formic acid. (b) Second purification with water/0.1% TFA and MeCN/0.1% TFA.



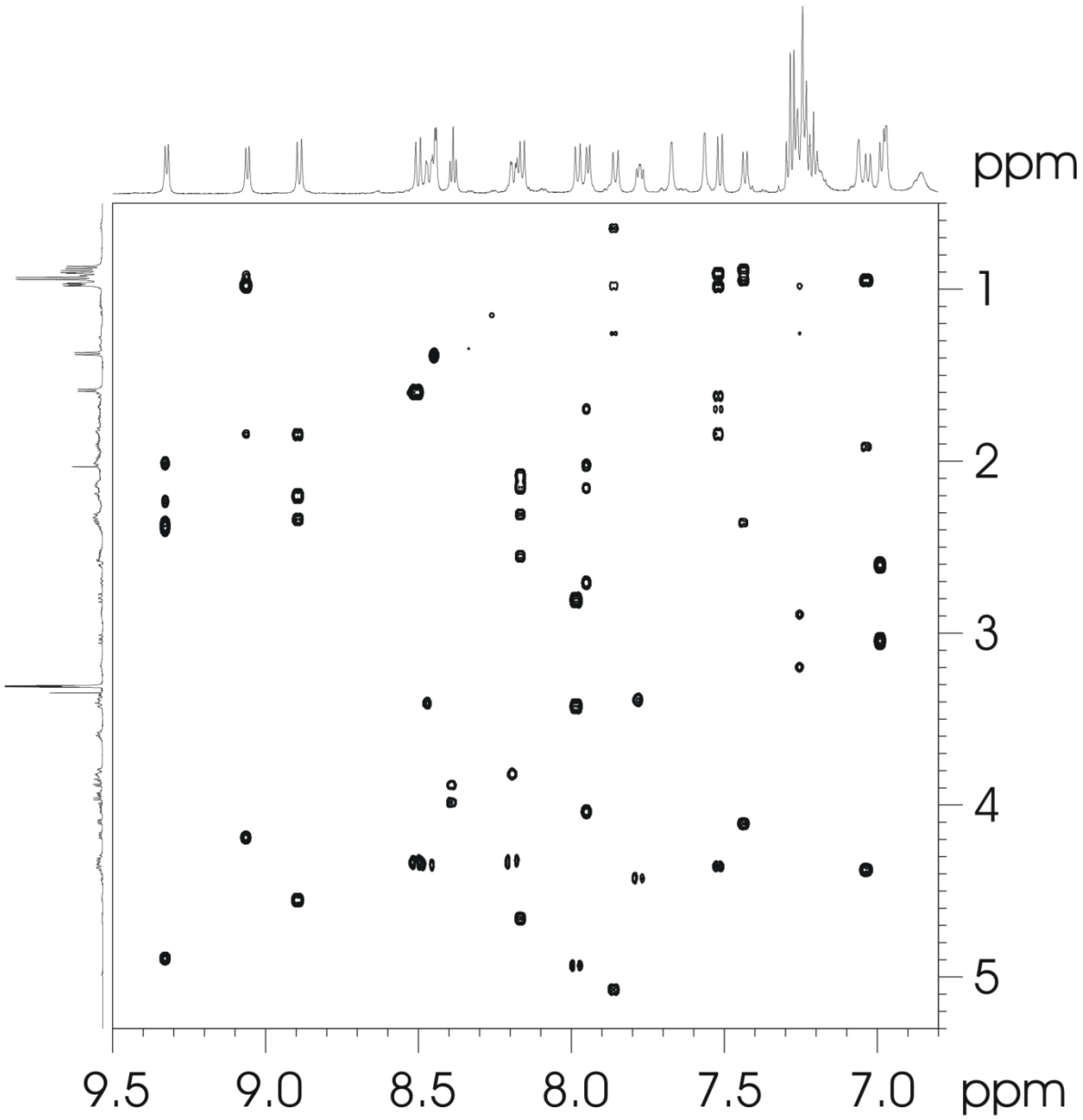
Supplementary Figure S2. Chromatograms of the HPLC purification of caulosegnin II. (a) First purification with water/0.05% formic acid and MeOH/0.045% formic acid. (b) Second purification with water/0.1% TFA and MeCN/0.1% TFA.



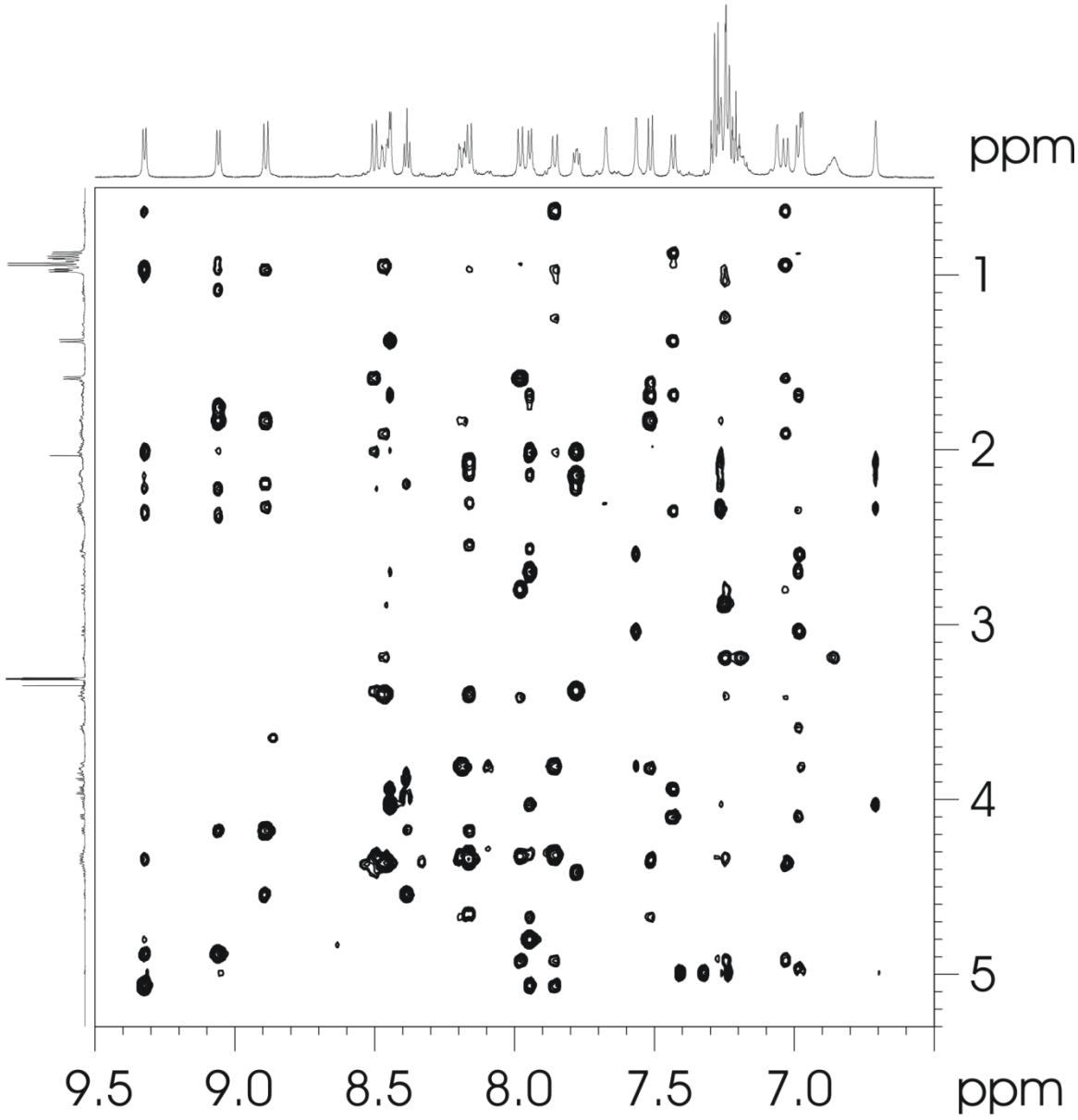
Supplementary Figure S3. Chromatograms of the HPLC purification of caulosegnin III. (a) First purification with water/0.05% formic acid and MeOH/0.045% formic acid. (b) Second purification with water/0.1% TFA and MeCN/0.1% TFA.



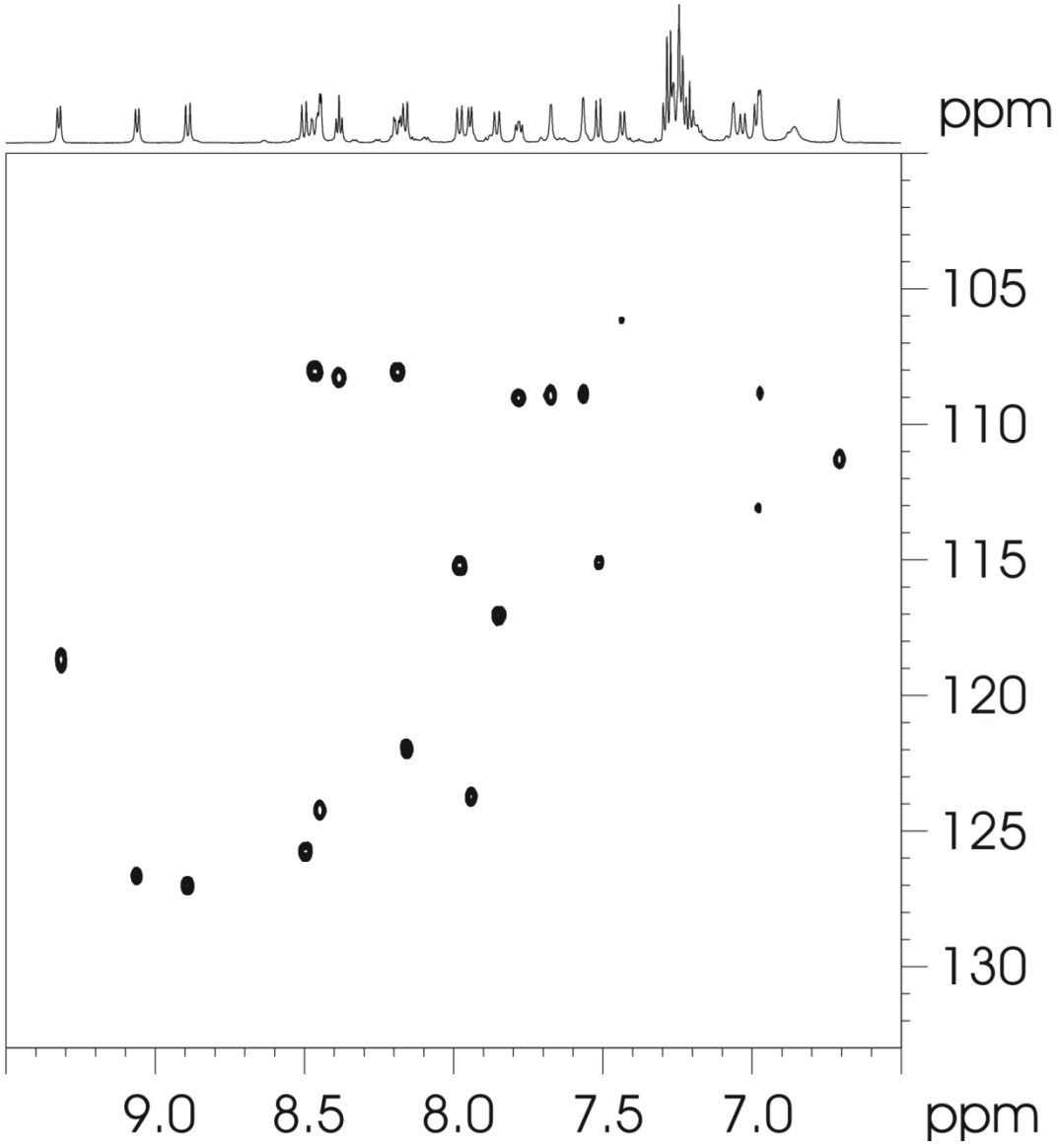
Supplementary Figure S4. DQF-COSY spectrum in the fingerprint region of caulosegnin I in methanol-d_3 at 288 K.



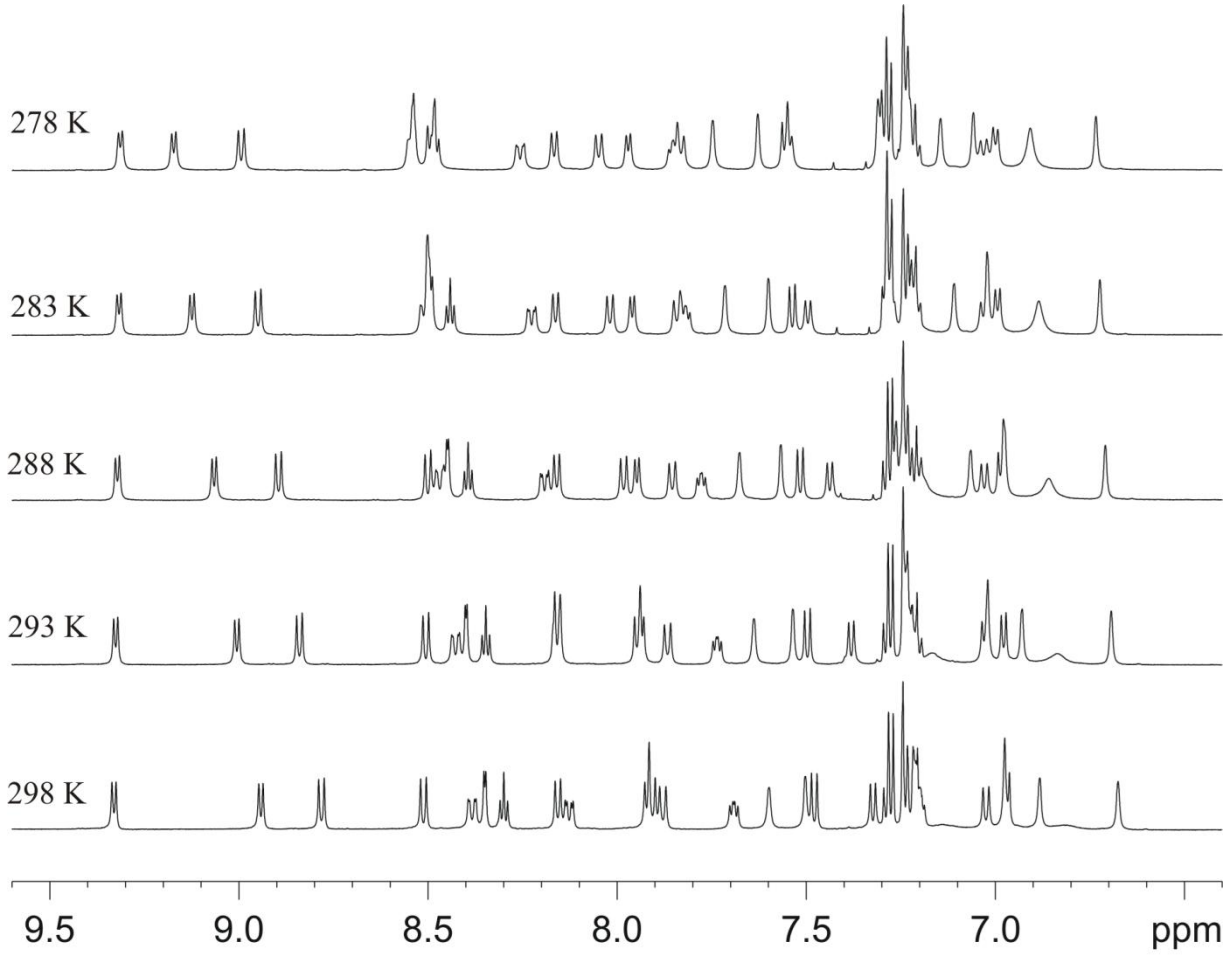
Supplementary Figure S5. Section of the TOCSY spectrum (mixing time 140 ms) of caulosegnin I in methanol-d_3 at 288 K.



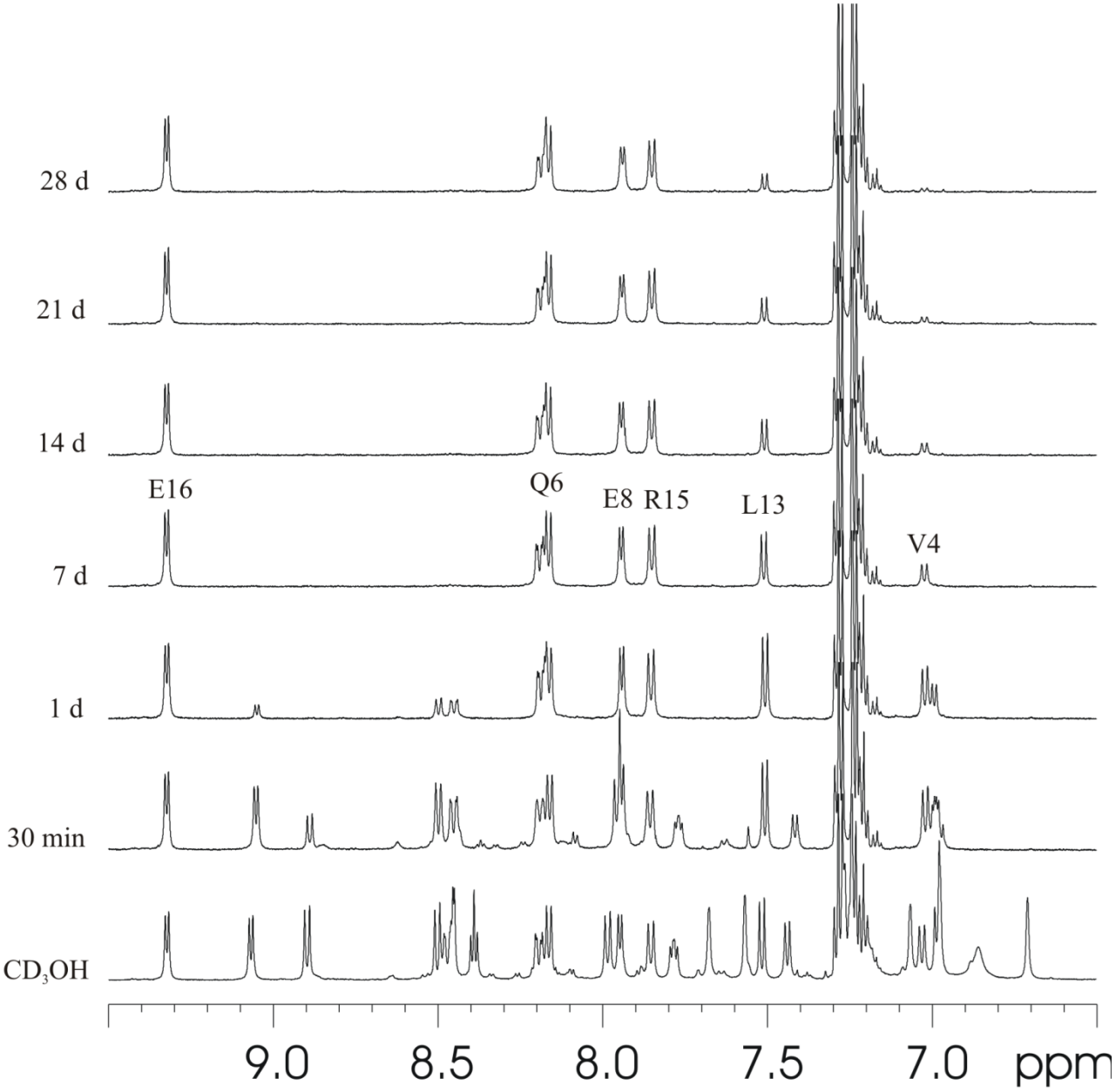
Supplementary Figure S6. Section of the NOESY spectrum (mixing time 300 ms) of caulosegnin I in methanol-d_3 at 288 K.



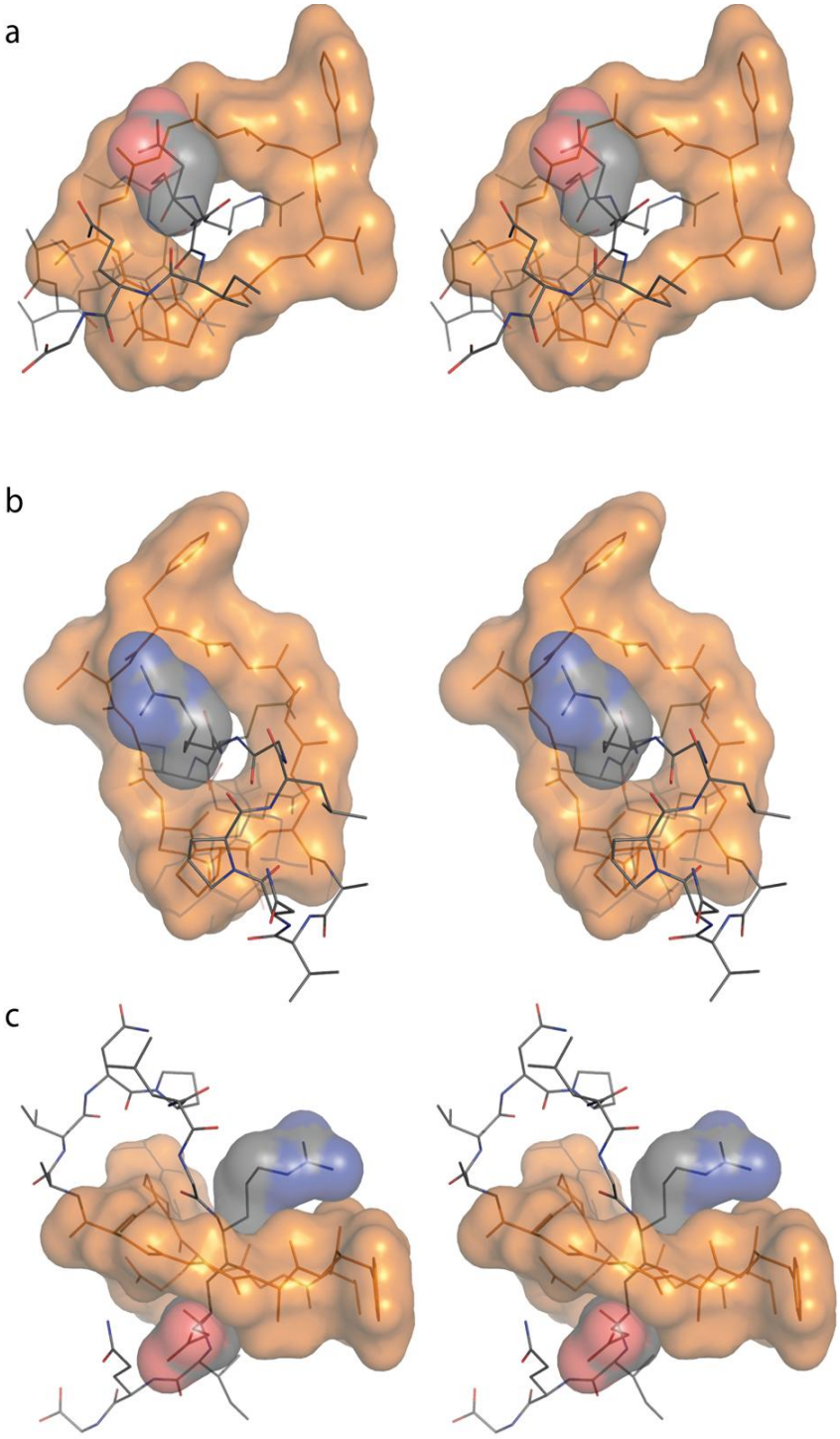
Supplementary Figure S7. ^{1}H, ^{15}NHSQC spectrum of caulosegnin I in methanol-d_3 at 288 K.



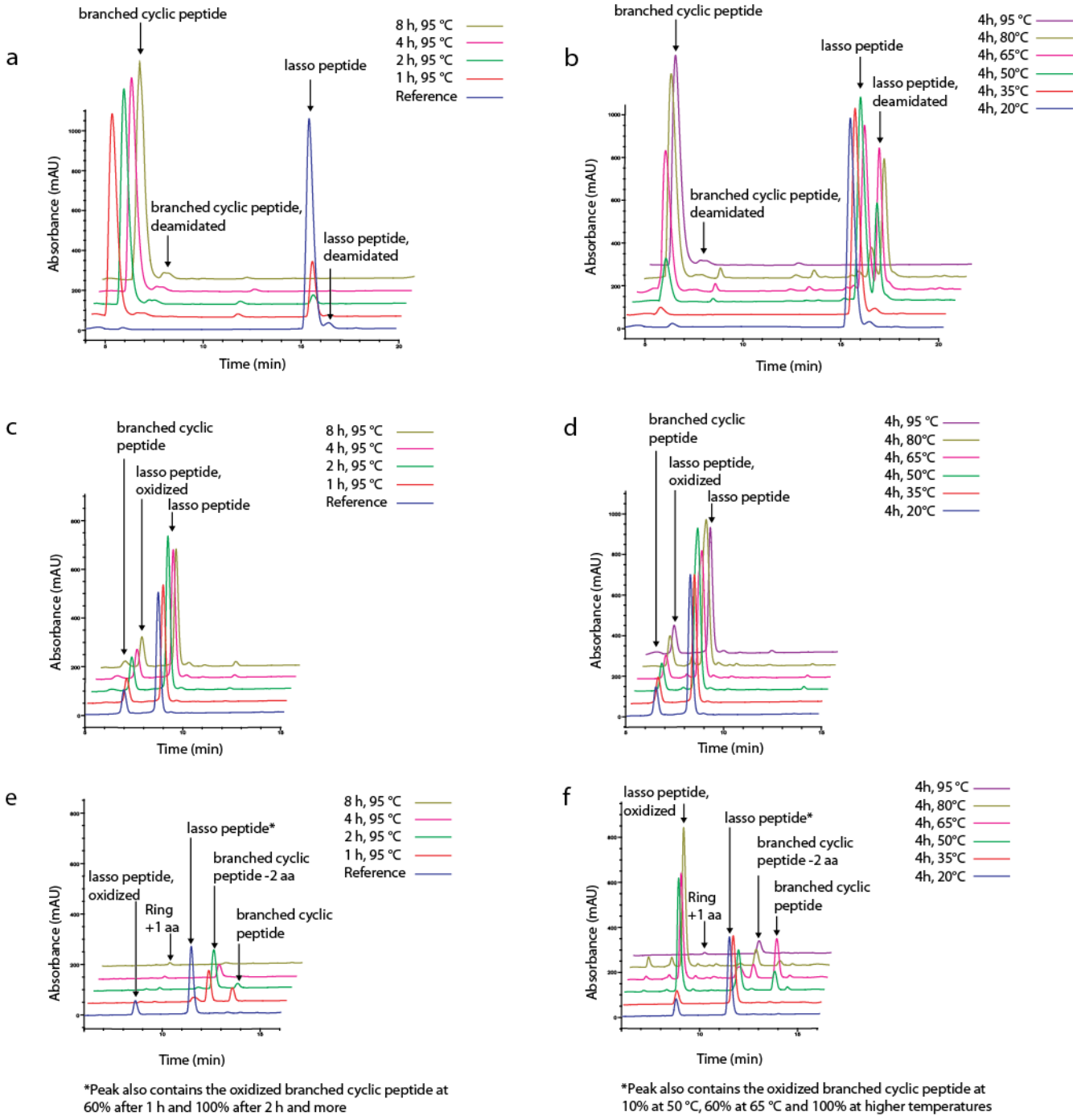
Supplementary Figure S8. ^{1}H variable temperature spectra in NH region of cauloesegnin I in methanol-d_3.



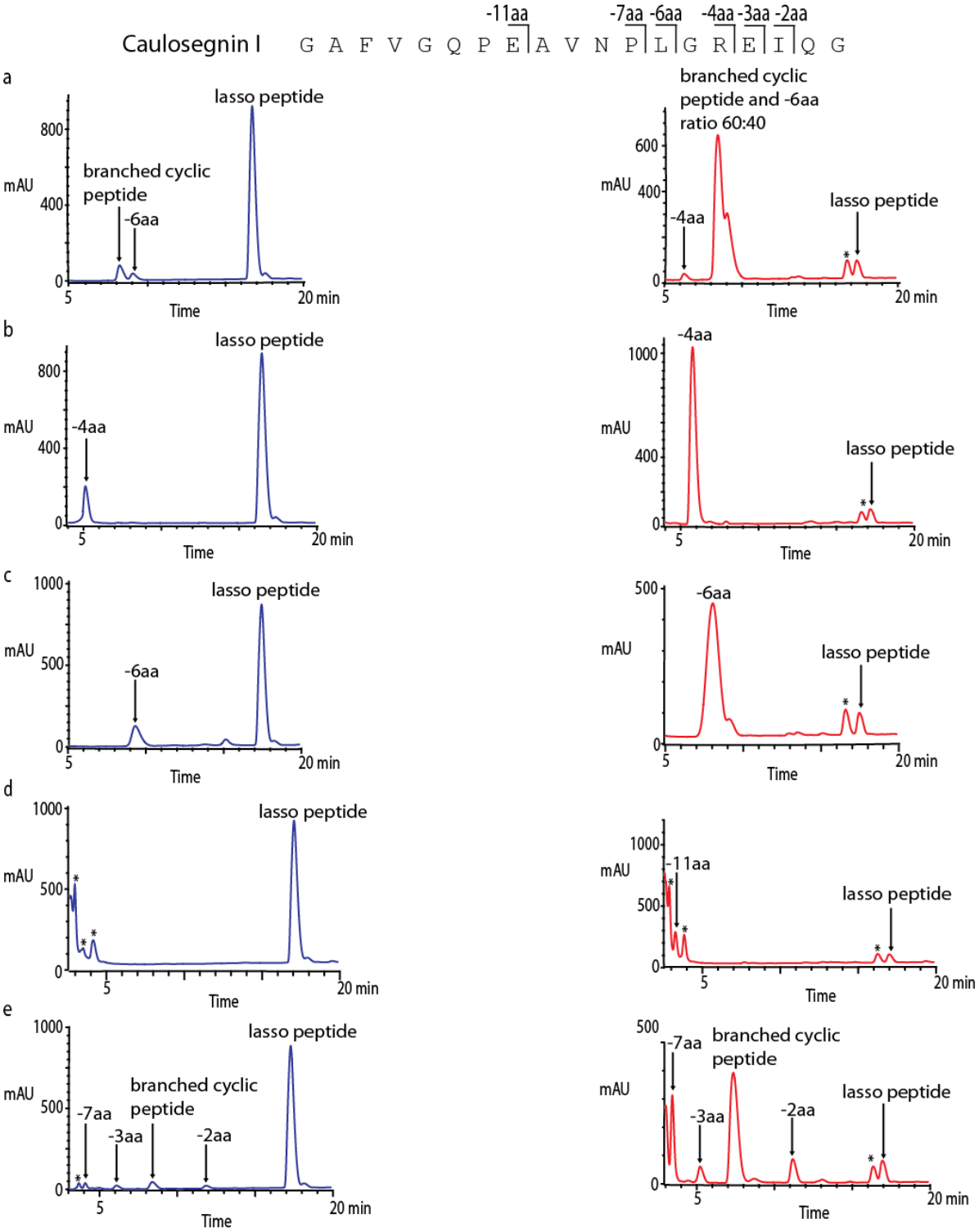
Supplementary Figure S9. ^{1}H variable delay spectra in NH region of caulosegnin I in methanol-d_4. Spectrum in methanol-d_3 is shown as reference. The spectra were recorded at 288 K.



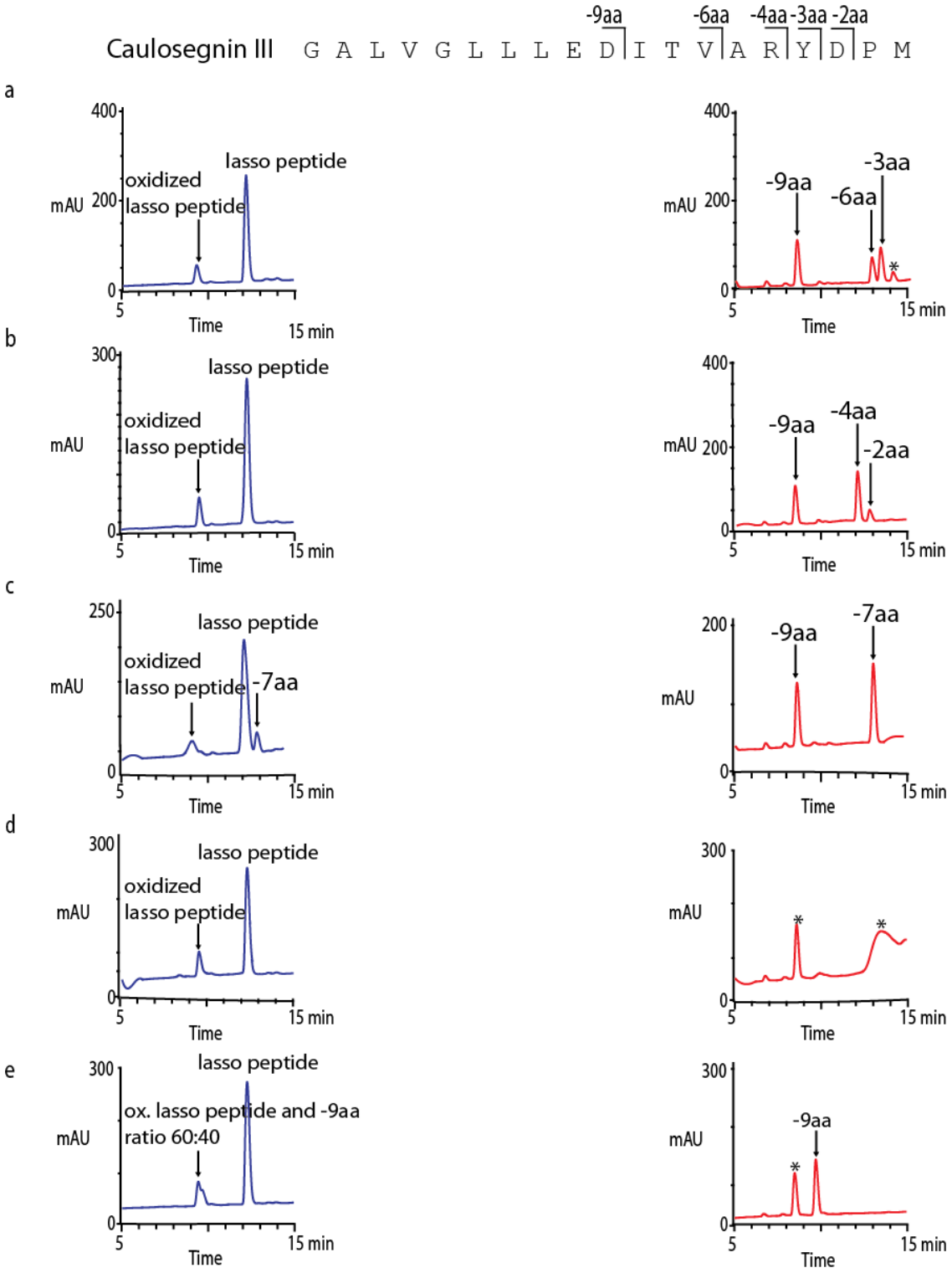
Supplementary Figure S10. Interactions between the C-terminal tail and the macrocyclic ring of caulosegnin I shown in stereoview. The surface of the ring is colored in orange, the surface of the side chains from Arg15 and Glu16 are colored by elements. Steric hindrance caused by the depicted side chains traps the tail within the ring. (a) Shown from below, (b) from the side and (c) from above the ring.



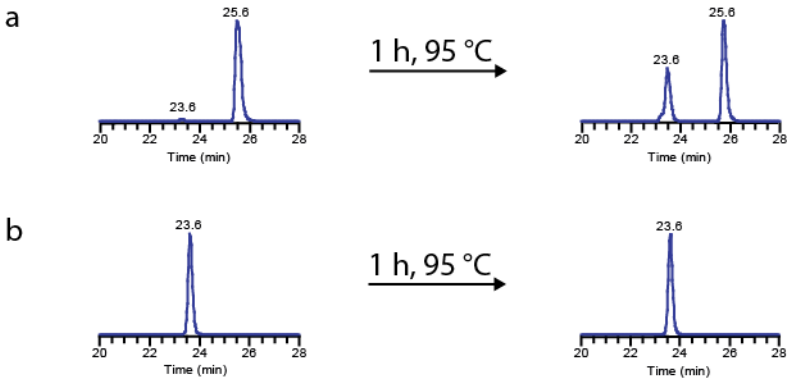
Supplementary Figure S11. Chromatograms of the heat stability assays of caulosegnin IIII. (a) Incubation of caulosegnin I at 95 °C for different time intervals. (b) Incubation of caulosegnin I at different temperatures for 4 h. (c) Incubation of caulosegnin II at 95 °C for different time intervals. (d) Incubation of caulosegnin II at different temperatures for 4 h. (e) Incubation of caulosegnin III at 95 °C for different time intervals. (f) Incubation of caulosegnin III at different temperatures for 4 h.



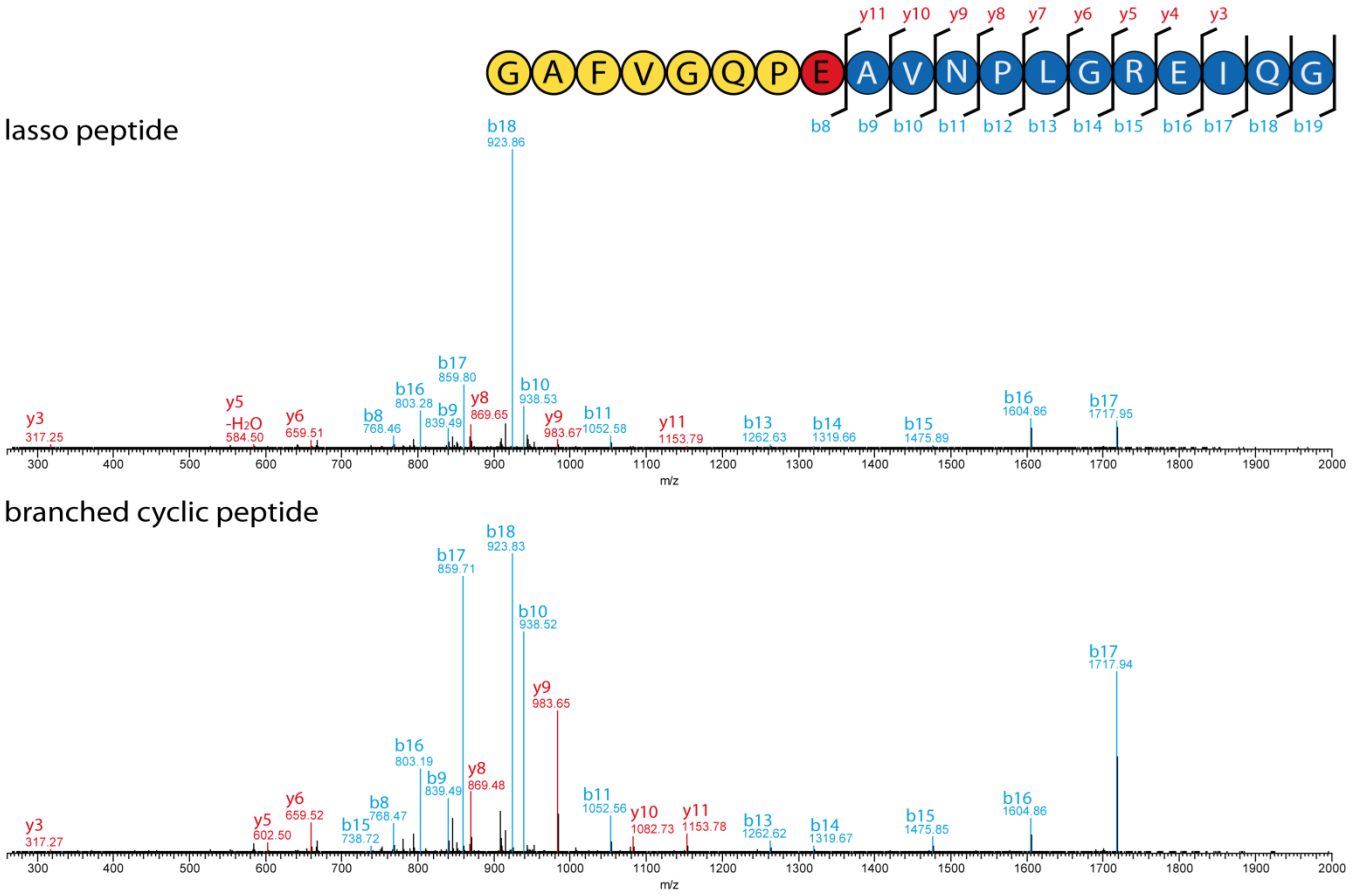
Supplementary Figure S12. Chromatograms of the protease stability assays of caulosegnin I. The digests of lasso peptide is shown in blue, the digest of thermally unthreaded caulosegnin I ist shown in red. (a) Chymotrypsin, 4 h, 25 °C. (b) Trypsin, 4 h, 25 °C. (c) Elastase, 4 h, 37 °C. (d) Proteinase K, 4 h, 37 °C. (e) Carboxypeptidase Y, 16 h, 25 °C. Signals marked with an asterisk could not be assigned to a specific fragment.



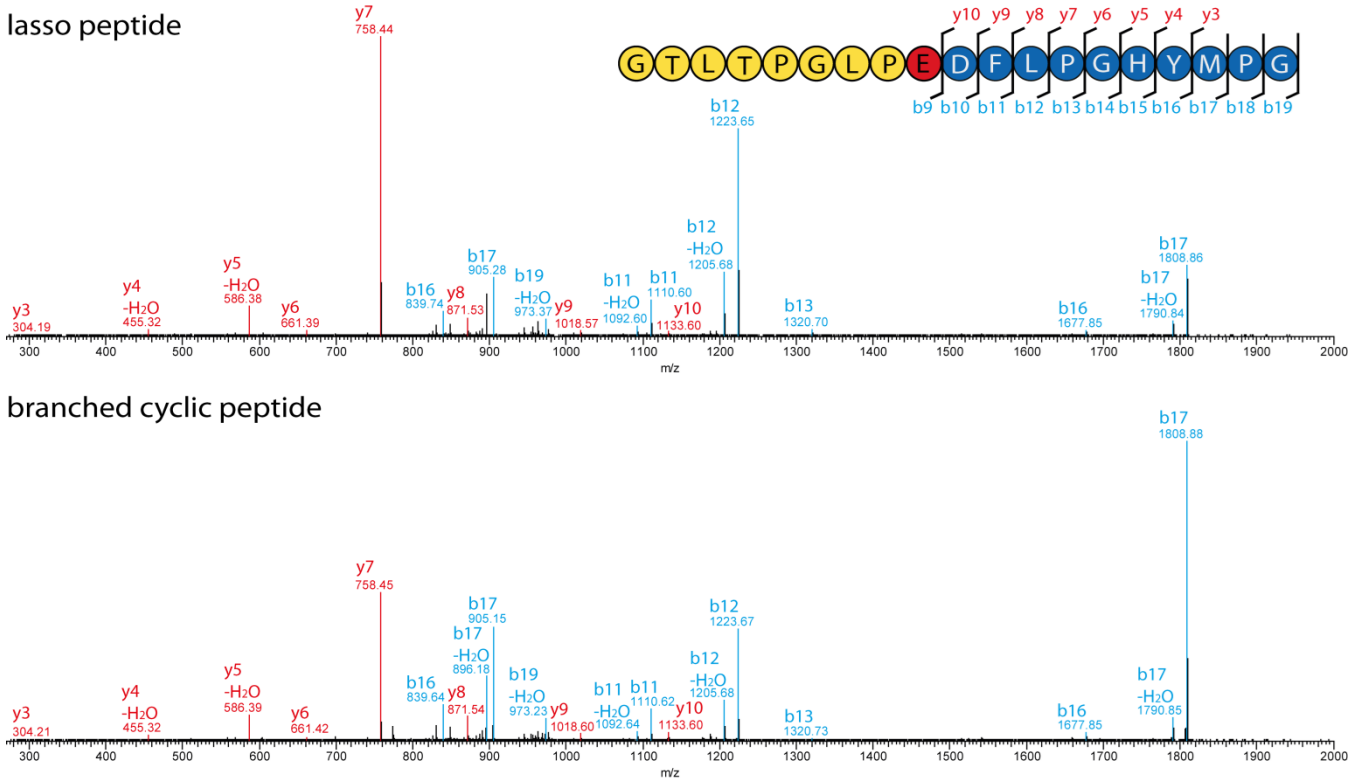
Supplementary Figure S13. Chromatograms of the protease stability assays of caulosegnin III. The digests of lasso peptide is shown in blue, the digest of thermally unthreaded caulosegnin I is shown in red. (a) Chymotrypsin, 4 h, 25 °C. (b) Trypsin, 4 h, 25 °C. (c) Elastase, 4 h, 37 °C. (d) Proteinase K, 4 h, 37 °C. (e) Carboxypeptidase Y, 16 h, 25 °C. Signals marked with an asterisk could not be assigned to a specific fragment.



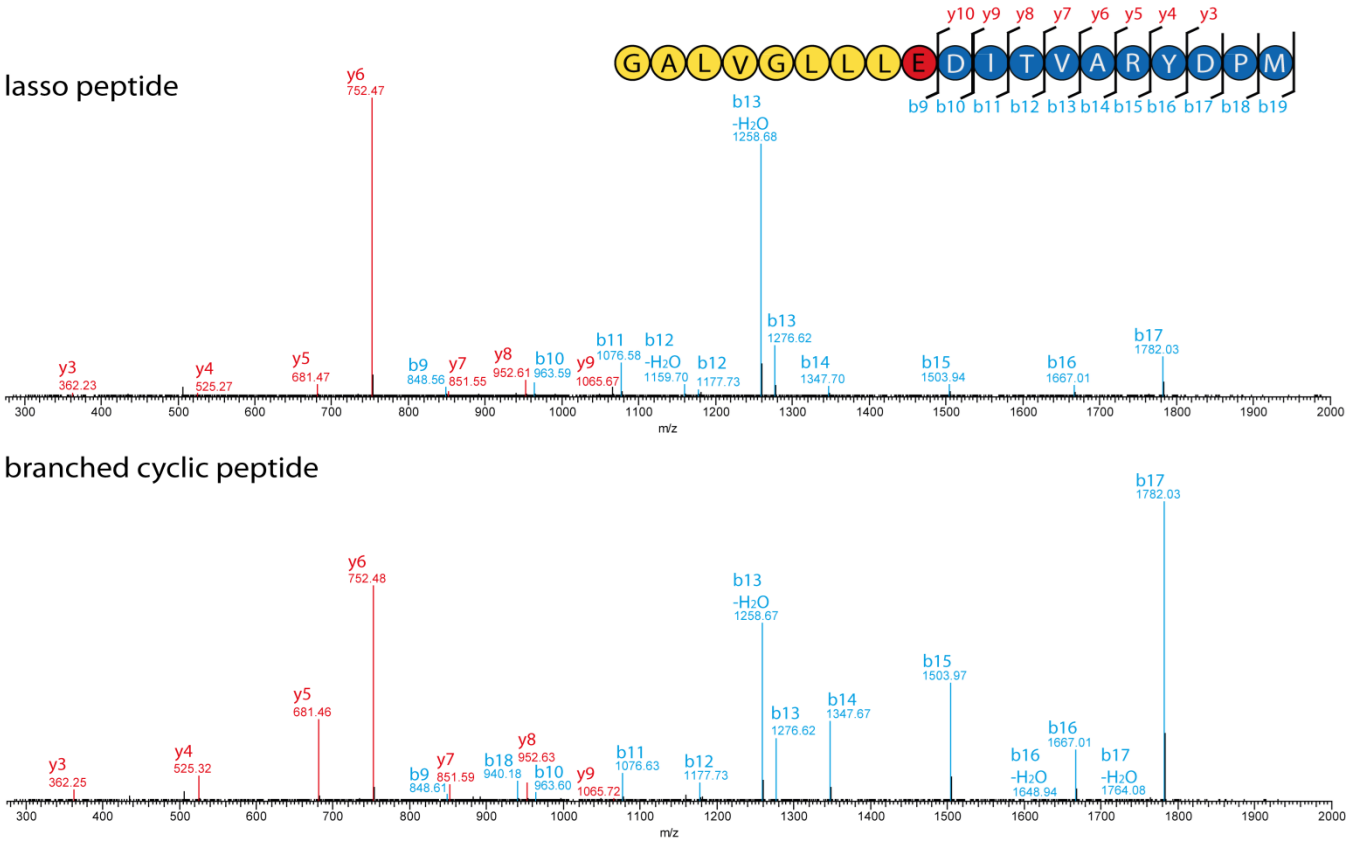
Supplementary Figure S14. Extracted ion chromatograms of (a) caulosegnin I P12A and (b) caulosegnin I E16A prior and after heating at 95 °C for 1 h. For caulosegnin I P12A the increase of the amount of unthreaded lasso peptide is observable, while heating does not affect caulosegnin I E16A at all. The lack of formation of a second peak would be due to the fact that the variant is either a heat stable lasso peptide or is already a branched cyclic peptide which can not further unthread. As it would make no sense that replacing the plug with a much smaller amino acid would lead to a heat stable variant, the latter must be the case. This behaviour was observed for every variant being designated as a branched cyclic peptide.



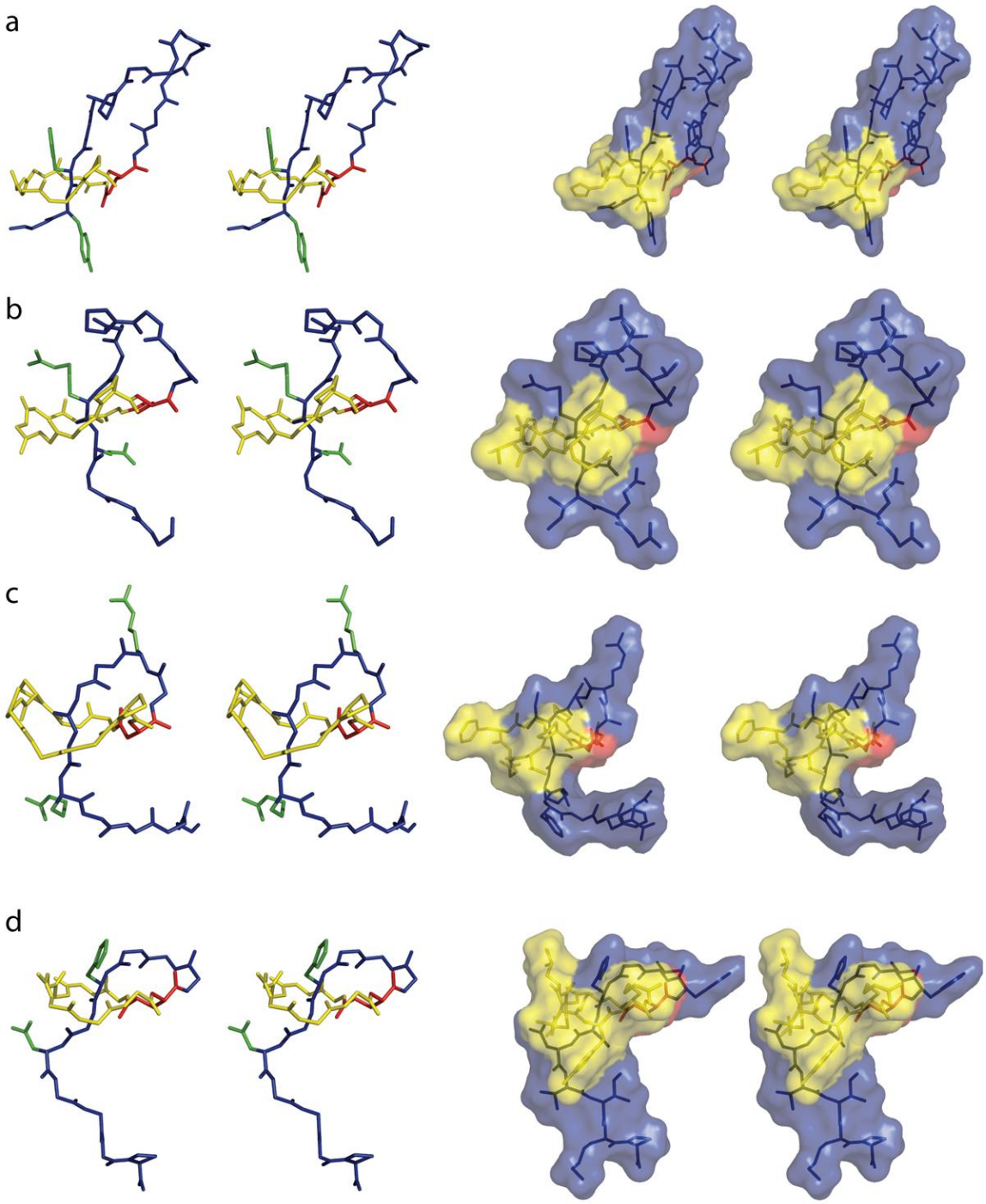
Supplementary Figure S15. MS^2 fragmentation of caulosegnin I and its branched cyclic analogue. Fragments of the b-series are highlighted in blue, fragments of the y-series in red.



Supplementary Figure S16. MS^2 fragmentation of caulosegnin II and its branched cyclic analogue. Fragments of the b-series are highlighted in blue, fragments of the y-series in red.



Supplementary Figure S17. MS^2 fragmentation of caulosegnin III and its branched cyclic analogue. Fragments of the b-series are highlighted in blue, fragments of the y-series in red.



Supplementary Figure S18. Comparison of class II lasso peptide structures. The ring forming aspartate/glutamate is shown in red, the remaining amino acids of the macrolactam ring are shown in yellow, the tail amino acids are shown in blue and the side chains of the amino acids which were proposed to sterically trap the tail inside the ring are highlighted in green. The shown lasso peptides (as sticks on the left and overlayed with surface maps on the right) are MccJ25[18] (a), caulosegnin I (b), capistruin[12] (c) and lariatin A[11] (d).

Supplementary Table S1. M9 Vitamin Mix.

component	amount
choline chloride	1.0 g
folic acid	1.0 g
pantothenic acid	1.0 g
nicotinamide	1.0 g
myo-inositol	2.0 g
pyridoxal hydrochloride	1.0 g
thiamine	1.0 g
riboflavin	0.1 g
disodium adenosine 5'-triphosphate	0.3 g
biotin	0.2 g
	ad 300 mL ddH_2O[a]

[a]10 M NaOH were added to the resulting suspension until everything was dissolved. Afterwards solution was sterile filtered and stored at 4 °C until needed.

Supplementary Table S2. Cloning primers. Introduced restriction sites are underlined.

name	sequence
Cseg_A1_NdeI_FP[a]	GGATTAC CAT ATG ACC AAG AAG AAC GCC ACC CAA G
Cseg_C_HindIII_RP[a]	ATAT AAG CTT TCA ACG GGT GTT CGC CTT CCT CG
LambdaTerm_McjProm_KpnI_FP[b]	ATAT GGT ACC AGC TTA ATT AGC TGA GCT TGG ACT CCT GTT GAT AG
LambdaTerm_McjProm_EcoRI_RP[b]	TATAT GAA TTC CAC TCC CTC CTG TAG TTA TGC CTT TAC ACT CAA TTG G
CsegBC-pET41a_EcoRI_FP[c]	ATATA GAA TTC ATG GAC CTG TGG CTC AGC GCC GGC GTG TAC
CsegA1-pET41a_KpnI_RP[c]	ATAT GGT ACC TCA GCC TTG GAT TTC GCG GCC GAG CGG ATT G

[a]*C. segnis* chromosomal DNA was used as template
[b]pJP3 was used as template (Pan, S. J.; Cheung, W. L.; Link, A. J. *Prot. Expr. Purif.* **2010**, *71*, 200-206).
[c]Cseg AAABC pET41a was used as template

Supplementary Table S3a. Caulosegnin I mutagenesis primers.

All primers were 5'-phosphorylated. Mutated sequences are marked in bold. A mutant was generated via PCR with a FP or RP carrying the desired mutation and a respective WT RP or FP. CsegA1_T-P_BC pET41a was used as template for all PCRs.

name	sequence
CsegA1_WT_FP	AAT CCG CTC GGC CGC GAA ATC CAA GGC TGA GGT ACC AGC TTA ATT AGC TGA GCT TGG ACT CCT GTT G
CsegA1_WT_RP	GAC CGC TTC GGG CTG GCC GAC GAA CGC GCC TTG CGT CAG GCG ATG GGC GTC GCC GAC GC
CsegA1_T-2A_RP	GAC CGC TTC GGG CTG GCC GAC GAA CGC GCC TTG **CGC** CAG GCG ATG GGC GTC GCC GAC GC
CsegA1_T-2C_RP	GAC CGC TTC GGG CTG GCC GAC GAA CGC GCC TTG **GCA** CAG GCG ATG GGC GTC GCC GAC GC
CsegA1_T-2I_RP	GAC CGC TTC GGG CTG GCC GAC GAA CGC GCC TTG **AAT** CAG GCG ATG GGC GTC GCC GAC GC
CsegA1_T-2S_RP	GAC CGC TTC GGG CTG GCC GAC GAA CGC GCC TTG **GCT** CAG GCG ATG GGC GTC GCC GAC GC
CsegA1_T-2V_RP	GAC CGC TTC GGG CTG GCC GAC GAA CGC GCC TTG **AAC** CAG GCG ATG GGC GTC GCC GAC GC
CsegA1_G1A_RP	GAC CGC TTC GGG CTG GCC GAC GAA CGC **GGC** TTG CGT CAG GCG ATG GGC GTC GCC GAC GC
CsegA1_G1C_RP	GAC CGC TTC GGG CTG GCC GAC GAA CGC **GCA** TTG CGT CAG GCG ATG GGC GTC GCC GAC GC
CsegA1_G1F_RP	GAC CGC TTC GGG CTG GCC GAC GAA CGC **AAA** TTG CGT CAG GCG ATG GGC GTC GCC GAC GC
CsegA1_E8D_RP	GAC CGC **ATC** GGG CTG GCC GAC GAA CGC GCC TTG CGT CAG GCG ATG GGC GTC GCC GAC GC
CsegA1_P12A_FP	AAT **GCG** CTC GGC CGC GAA ATC CAA GGC TGA GGT ACC AGC TTA ATT AGC TGA GCT TGG ACT CCT GTT G
CsegA1_ΔL13_FP	AAT CCG GGC CGC GAA ATC CAA GGC TGA GGT ACC AGC TTA ATT AGC TGA GCT TGG ACT CCT GTT G
CsegA1_R15A_FP	AAT CCG CTC GGC **GCG** GAA ATC CAA GGC TGA GGT ACC AGC TTA ATT AGC TGA GCT TGG ACT CCT GTT G
CsegA1_E16A_FP	AAT CCG CTC GGC CGC **GCG** ATC CAA GGC TGA GGT ACC AGC TTA ATT AGC TGA GCT TGG ACT CCT GTT G
CsegA1_E16D_FP	AAT CCG CTC GGC CGC **GAT** ATC CAA GGC TGA GGT ACC AGC TTA ATT AGC TGA GCT TGG ACT CCT GTT G
CsegA1_E16K_FP	AAT CCG CTC GGC CGC **AAA** ATC CAA GGC TGA GGT ACC AGC TTA ATT AGC TGA GCT TGG ACT CCT GTT G

Supplementary Table S3b. Caulosegnin I mutagenesis primers.

All primers were 5'-phosphorylated. Mutated sequences are marked in bold. A mutant was generated via PCR with a FP or RP carrying the desired mutation and a respective WT RP or FP. CsegA1_T-P_BC pET41a was used as template for all PCRs.

name	sequence
CsegA1_E16L_FP	AAT CCG CTC GGC CGC **CTG** ATC CAA GGC TGA GGT ACC AGC TTA ATT AGC TGA GCT TGG ACT CCT GTT G
CsegA1_E16M_FP	AAT CCG CTC GGC CGC **ATG** ATC CAA GGC TGA GGT ACC AGC TTA ATT AGC TGA GCT TGG ACT CCT GTT G
CsegA1_E16N_FP	AAT CCG CTC GGC CGC **AAC** ATC CAA GGC TGA GGT ACC AGC TTA ATT AGC TGA GCT TGG ACT CCT GTT G
CsegA1_E16Q_FP	AAT CCG CTC GGC CGC **CAG** ATC CAA GGC TGA GGT ACC AGC TTA ATT AGC TGA GCT TGG ACT CCT GTT G
CsegA1_E16A_I17A_FP	AAT CCG CTC GGC CGC **GCG GCC** CAA GGC TGA GGT ACC AGC TTA ATT AGC TGA GCT TGG ACT CCT GTT G
CsegA1_I17A_FP	AAT CCG CTC GGC CGC GAA **GCG** CAA GGC TGA GGT ACC AGC TTA ATT AGC TGA GCT TGG ACT CCT GTT G
CsegA1_I17STOP_FP	AAT CCG CTC GGC CGC GAA **TAA** CAA GGC TGA GGT ACC AGC TTA ATT AGC TGA GCT TGG ACT CCT GTT G
CsegA1_Q18STOP_FP	AAT CCG CTC GGC CGC GAA ATC **TAA** GGC TGA GGT ACC AGC TTA ATT AGC TGA GCT TGG ACT CCT GTT G
CsegA1_G19STOP_FP	AAT CCG CTC GGC CGC GAA ATC CAA **TAA** TGA GGT ACC AGC TTA ATT AGC TGA GCT TGG ACT CCT GTT G

Supplementary Table S4. Caulosegnin II mutagenesis primers.

All primers were 5'-phosphorylated. Mutated sequences are marked in bold. A mutant was generated via PCR with a FP carrying the desired mutation and a WT RP. CsegA1-QG-A2_T-P_BC pET41a was used as template for all PCRs.

name	sequence
CsegA2_WT_RP	ATC CTC CGG CAG GCC CGG AGT GAG CGT GCC TTG CGT CAG GCG ATG GGC GTC GCC GAC GC
CsegA2_P13A_FP	TTC CTG **GCG** GGC CAC TAC ATG CCG GGC TGA GGT ACC AGC TTA ATT AGC TGA GCT TGG ACT CCT GTT G
CsegA2_H15A_FP	TTC CTG CCT GGC **GCG** TAC ATG CCG GGC TGA GGT ACC AGC TTA ATT AGC TGA GCT TGG ACT CCT GTT G
CsegA2_Y16A_FP	TTC CTG CCT GGC CAC **GCG** ATG CCG GGC TGA GGT ACC AGC TTA ATT AGC TGA GCT TGG ACT CCT GTT G
CsegA2_M17A_FP	TTC CTG CCT GGC CAC TAC **GCG** CCG GGC TGA GGT ACC AGC TTA ATT AGC TGA GCT TGG ACT CCT GTT G
CsegA2_P18A_FP	TTC CTG CCT GGC CAC TAC ATG **GCG** GGC TGA GGT ACC AGC TTA ATT AGC TGA GCT TGG ACT CCT GTT G

Supplementary Table S5. Caulosegnin III mutagenesis primers.

All primers were 5'-phosphorylated. Mutated sequences are marked in bold. A mutant was generated via PCR with a FP carrying the desired mutation and a WT RP. CsegA1-QG-A3_T-P_BC pET41a was used as template for all PCRs.

name	sequence
CsegA3_WT_RP	GTC CTC GAG CAG CAG GCC CAC CAG CGC GCC TTG CGT CAG GCG ATG GGC GTC GCC GAC GC
CsegA3_V13P_FP	ATC ACC **CCG** GCC CGC TAC GAC CCG ATG TGA GGT ACC AGC TTA ATT AGC TGA GCT TGG ACT CCT GTT G
CsegA3_R15A_FP	ATC ACC GTG GCC **GCG** TAC GAC CCG ATG TGA GGT ACC AGC TTA ATT AGC TGA GCT TGG ACT CCT GTT G
CsegA3_Y16A_FP	ATC ACC GTG GCC CGC **GCG** GAC CCG ATG TGA GGT ACC AGC TTA ATT AGC TGA GCT TGG ACT CCT GTT G
CsegA3_D17A_FP	ATC ACC GTG GCC CGC TAC **GCG** CCG ATG TGA GGT ACC AGC TTA ATT AGC TGA GCT TGG ACT CCT GTT G
CsegA3_P18A_FP	ATC ACC GTG GCC CGC TAC GAC **GCG** ATG TGA GGT ACC AGC TTA ATT AGC TGA GCT TGG ACT CCT GTT G
CsegA3_M19A_FP	ATC ACC GTG GCC CGC TAC GAC CCG **GCG** TGA GGT ACC AGC TTA ATT AGC TGA GCT TGG ACT CCT GTT G

Supplementary Table S6. Primers for modification of plasmid via deletion, insertion or replacement of DNA sequences. All primers were 5'-phosphorylated. Inserted sequences are underlined.

name	sequence
CsegA1BC_FP[a]	TGA GAC GGC GGC CTT GGC CCC GAG CGC CGG G
CsegA1BC_RP[a]	GCC TTG GAT TTC GCG GCC GAG CGG ATT GAC CGC TTC GG
CsegA1A2BC_FP[a]	TGA GAC GGC GGC CTT GGC CCC GAG CGC CGG G
CsegA1A2BC_RP[a]	GCC CGG CAT GTA GTG GCC AGG CAG GAA ATC CTC CGG CAG G
CsegA2BC_FP[b]	ATG ACC AAG ACC CAT CGC CTG ATC CGC CTG GGC GAC GCC C
CsegA2BC_RP[b]	ATG TAT ATC TCC TTC TTA AAG TTA AAC AAA ATT ATT TCT AGA GGG GAA TTG TTA TCC GCT C
CsegA3BC_FP[a]	ATG ACG TCT AGG TTT CAA CTG CTG CGA TTG GGC AAG GCC
CsegA3BC_RP[a]	ATG TAT ATC TCC TTC TTA AAG TTA AAC AAA ATT ATT TCT AGA GGG GAA TTG TTA TCC GCT C
CsegA1-QG-A2BC_FP[b]	GGC ACG CTC ACT CCG GGC CTG CCG GAG GAT TTC CTG CC
CsegA1-QG-A2BC_RP[b]	TTG CGT CAG GCG ATG GGC GTC GCC GAC GCG C
CsegA1-QG-A3BC_FP[a]	GGC GCG CTG GTG GGC CTG CTG CTC GAG GAC ATC AC
CsegA1-QG-A3BC_RP[a]	TTG CGT CAG GCG ATG GGC GTC GCC GAC GCG C
CsegA1_RBS_FP[a]	<u>AGAGGAGAAATTAACC</u> ATG GAC CTG TGG CTC AGC GCC GGC GTG TAC G
CsegA1_RBS_RP[a]	TCA GCC TTG GAT TTC GCG GCC GAG CGG ATT GAC
CsegA1-QG-A2-T-P-BC FP[c]	<u>TTC CTG CCT GGC CAC TAC ATG CCG GGC</u> TGA GGT ACC AGC TTA ATT AGC TGA GCT TGG ACT CCT GTT G
CsegA1-QG-A2-T-P-BC RP[c]	<u>ATC CTC CGG CAG GCC CGG AGT GAG CGT</u> GCC TTG CGT CAG GCG ATG GGC GTC GCC GAC GC
CsegA1-QG-A3-T-P-BC FP[c]	<u>ATC ACC GTG GCC CGC TAC GAC CCG ATG</u> TGA GGT ACC AGC TTA ATT AGC TGA GCT TGG ACT CCT GTT G
CsegA1-QG-A3-T-P-BC RP[c]	<u>GTC CTC GAG CAG CAG GCC CAC CAG CGC</u> GCC TTG CGT CAG GCG ATG GGC GTC GCC GAC GC

[a]Cseg AAABC pET41a was used as template
[b]Cseg A1A2BC pET41a was used as template
[c]Cseg A1_T-P_BC pET41a was used as template

Supplementary Table S7. Production of different constructs used for optimization of caulosegin IIII production. Presented absolute values are the integrals of the UV peaks of the lasso peptides from LCMS measurements of 100 µL of the respective extracts.

name	UV peak area test 1	UV peak area test 2	UV peak area test 3	UV peak area test 4	mean value
cseg *AAABC* pET41a[a]	2534249	2898494	2124794	2560477	2529504
cseg *A1BC* pET41a	47671898	32481826	47204768	28266509	38906251
cseg *A1*_RBS_*BC* pET41a	406593390	454829558	377863698	426010521	416324292
cseg *A1*_TP_*BC* pET41a	518756822	433002760	719951408	885837566	639387139
cseg *A2BC* pET41a	2206210	1839511	1411016	2252130	1927217
cseg *A1-QG-A2BC* pET41a	10840516	18858142	6480280	16528542	13176870
cseg *A1-QG-A2*_TP_*BC* pET41a	239859039	403669011	308412837	336169281	322027542
cseg *A3BC* pET41a	1985625	2510361	233457	1419537	1537245
cseg *A1-QG-A3BC* pET41a	5242645	4489823	3627624	3917621	4319428
Cseg *A1-QG-A3*_TP_*BC* pET41a	276867510	156550224	161126313	273336172	216970055

[a]Only caulosegnin I was produced in amounts big enough to yield an UV signal and thus was the only lasso peptide which production could be determined by UV peak integration.

Supplementary Table S8. Overview of the sensitivity against proteolytic degradation of the caulosegnins and their branched cyclic analogues. If more than one degradation product was detected, the main product is shown in bold. (see also figure S5-S6)

	trypsin	chymotrypsin	elastase	proteinase K	carboxy-peptidase Y
caulosegnin I	full length, traces of -4 aa	full length, traces of -6 aa	full length, traces of -6 aa	only full length	full length, traces of -2 aa, -3 aa, **-7 aa**
caulosegnin I, unthreaded	**-4 aa**[a]	-4 aa, **-6 aa**[a]	**-6 aa**[a]	**-11 aa**[a]	-2 aa, -3 aa, **-7 aa**[a]
caulosegnin II	only full length	only full length	only full length	only full length	only full length
caulosegnin III	only full length	only full length	full length, traces of -7 aa	only full length	full length, traces of -9 aa
caulosegnin III, unthreaded	**-4 aa**, -9 aa, -2 aa	**-3 aa**, -6 aa, - 9aa	**-7 aa**, -9 aa	complete degradation	**-9 aa**

[a]Full length peptide was still detectable

Supplementary Table S9. Structural Statistics for the Family of 15 Structures Selected to Represent the Solution Structure of caulosegnin I.

Restraining Constraints	Constraints Violations
Total: 181	Distance violations, >0.5 Å: 0
distance, i=j: 74	RMS deviations: 0.02 Å
distance, |i-j|=1: 36	Dihedral violations, > 5°: 0
distance, |i-j|>1: 38	RMS deviation: 2.4°
dihedral: 33	Average pairwise RMS deviation (Thr^{2}-Phe^{18})
hydrogen bond: 0	Backbone atoms: 0.03 Å
Constraints/residue: 9.5	All heavy atoms: 0.35 Å

Supplementary Table S10. Assignment of ^{1}H and ^{15}N Signals (ppm) of caulosegnin I in methanol-d_3 at 288 K.

AA	NH	αH	βH	others	^{15}N
Gly1	7.778	4.415; 3.378	/	/	108.9
Ala2	8.501	4.324	1.589	/	125.7
Phe3	7.979	4.922	2.797; 3.419		115.1
Val4	7.031	4.367	1.905	γCH3: 0.937	114.8
Gly5	8.465	4.343; 3.396		/	108.0
Gln6	8.161	4.650	2.079; 2.318	γCH2: 2.138; 2.550	121.9
				εNH_2: 7.674; 7.061	εNH_2: 108.8
Pro7	/	4.802	2.562; 1.753	γCH2: 1.880; 2.135	/
				δCH2: 3.592; 3.847	
Glu8	7.945	4.029	1.683; 2.696	γCH2: 2.135; 2.013	123.7
Ala9	8.444	3.940	1.376	/	124.1
Val10	7.432	4.099	2.348	γCH3: 0.874; 0.936	106.2
Asn11	6.985	4.972	3.037; 2.594	δNH_2: 7.565; 6.971	113.0
					δNH_2: 108.7
Pro12	/	4.672	2.319; 1.934	γCH2: 1.996	/
				δCH2: 3.587; 3.821	
Leu13	7.514	4.345	1.832; 1.619	γCH: 1.682	115.1
				δCH3: 0.975; 0.894	
Gly14	8.188	4.315; 3.807		/	108.0
Arg15	7.855	5.061	0.635; 0.963	γCH2: 1.036; 1.241	117.0
				δCH2: 3.186; 2.885	
				εNH: 7.247	
Glu16	9.322	4.882	2.006; 2.226	γCH2: 2.366	118.8
Ile17	9.059	4.179	1.830	γCH2: 1.755; 1.081	126.6
				γCH3: 0.970	
				δCH3: 0.910	
Gln18	8.889	4.543	2.195; 1.836	γCH2: 2.318	127.0
				εNH_2: 7.263; 6.708	εNH_2: 111.1
Gly19	8.384	3.979; 3.870		/	108.3

Supplementary Table S11. Production of the different caulosegin I variants. Presented absolute values are the integrals of the UV peaks of the lasso peptides from LCMS measurements of 100 µL of the respective extracts.

name	UV peak area	relative production	retention time branched cyclic peptide / retention time lasso peptide
WT	389998258	1.00	24.5 min / 27.0 min
T2A	1318633	>0.01	24.5 min / 27.0 min
T2C	137451137	0.35	24.5 min / 27.0 min
T2I	11431440	0.03	24.5 min / 27.0 min
T2S	13727351	0.04	24.5 min / 27.0 min
T2V	34381092	0.09	24.5 min / 27.0 min
G1A	only mass	>0.01	23.8 min / 26.3 min
G1C	2560643	>0.01	24.5 min / 25.7 min
G1F	-	-	-
E8D	-	-	-
P12A	27804733	0.07	23.6 min / 25.6 min
ΔL13	15503862	0.04	20.1 min / 23.5 min
R15A	146869268	0.38	22.7 min / 24.4 min
E16A	3967317	0.01	23.6 min / -
E16D	300678905	0.77	22.7 min / 25.7 min
E16K	213060796	0.55	25.0 min / 26.7 min
E16L	42339595	0.11	25.8 min / 26.5 min
E16N	13363270[a]	0.03	22.4 min / 26.1 min
E16Q	95537464	0.25	22.5 min / 25.9 min
E16A/I17A	10207088	0.03	21.9 min / -
I17A	285451281	0.73	21.3 min / 24.5 min
I17STOP	only mass	>0.01	21.7 min / 24.7 min
Q18STOP	170273857	0.44	24.3 min / 27.1 min
G19STOP	423351745	1.09	23.2 min / 25.9 min

[a]UV peak area was integrated for the branched-cyclic analogue as it greatly exceeded the signal of the original lasso peptide.

Supplementary Table S12. Production of the different caulosegin II variants. Presented absolute values are the integrals of the UV peaks of the lasso peptides from LCMS measurements of 100 µL of the respective extracts.

name	UV peak area	relative production	retention time branched cyclic peptide / retention time lasso peptide
WT	251781008	1.00	26.0 min[a] / 27.5 min
P13A	104604208	0.42	26.1 min[a] / 26.9 min
H15A	126115859	0.50	25.4 min[a] / 26.6 min
Y16A	only mass	>0.01	25.6 min[b] / -
M17A	110407726	0.44	25.4 min[a] / 26.8 min
P18A	355436554	1.41	25.8 min[a] / 27.3 min

[a]Determined by incubating the extracts at 95 °C for 8 h to generate trace amounts of unthreaded lasso peptide.
[b]Retention time did not change after incubating the extract at 95 °C for 8 h.

Supplementary Table S13. Production of the different caulosegin III variants. Presented absolute values are the integrals of the UV peaks of the lasso peptides from LCMS measurements of 100 µL of the respective extracts.

name	UV peak area	relative production	retention time branched cyclic peptide / retention time lasso peptide
WT	135593658	1.00	29.5 min / 28.3 min
V13P	265951883	1.96	28.7 min / 27.8 min
R15A	-	-	-
Y16A	-	-	-
D17A	15218857	0.11	29.9 min / 29.2 min
P18A	65873784	0.49	29.7 min / 28.3 min
M19A	227569542	1.68	29.0 min / 27.5 min

Chemistry & Biology, Volume *20*

Supplemental Information

The Astexin-1 Lasso Peptides:

Biosynthesis, Stability, and Structural Studies

Marcel Zimmermann, Julian D. Hegemann, Xiulan Xie, and Mohamed A. Marahiel

Inventory of Supplemental Information

Supplemental Experimental Procedures

Strains and general methods

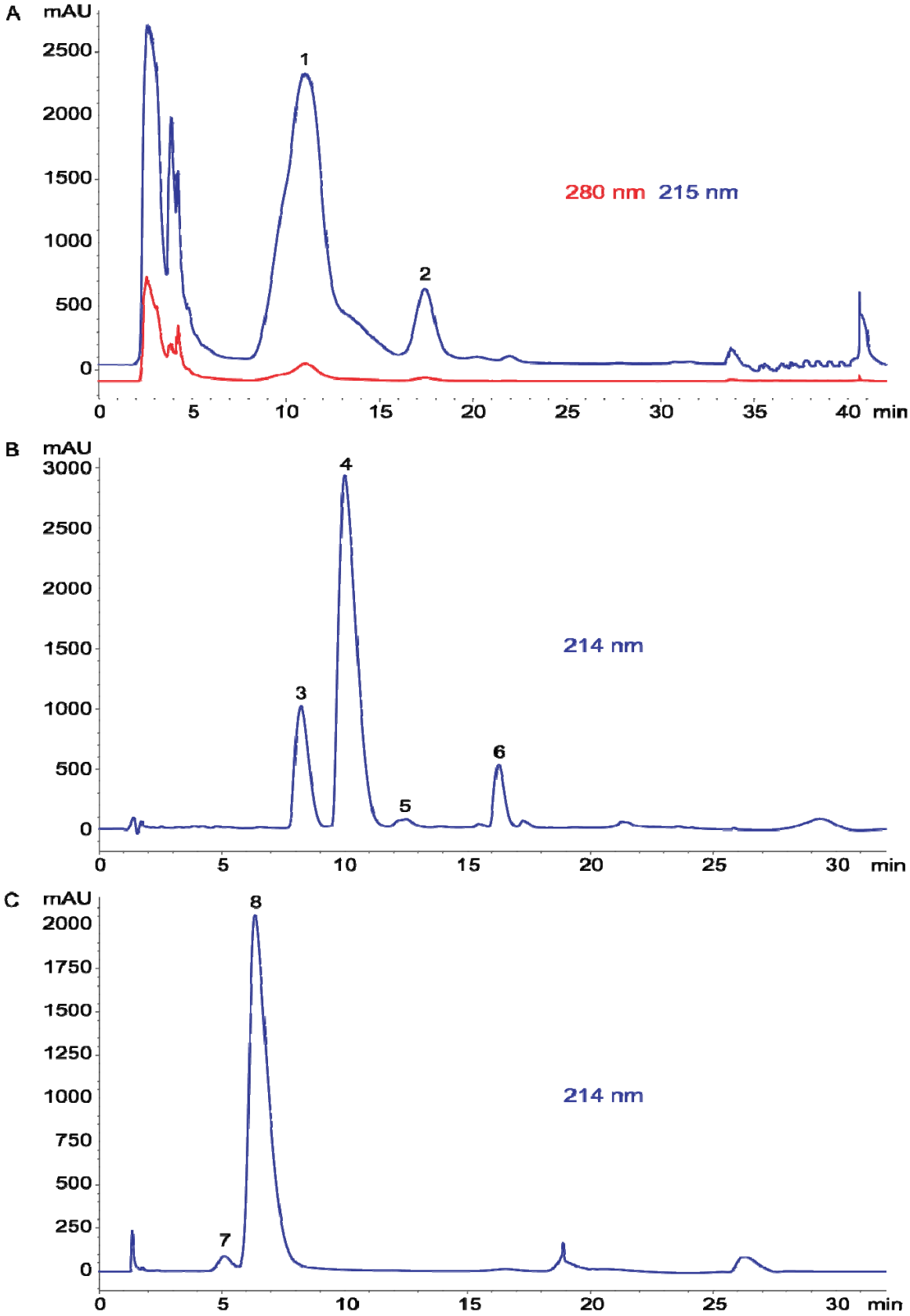
Cloning of the astexin-1 biosynthesis cluster

Construction of astexin-1 mutants

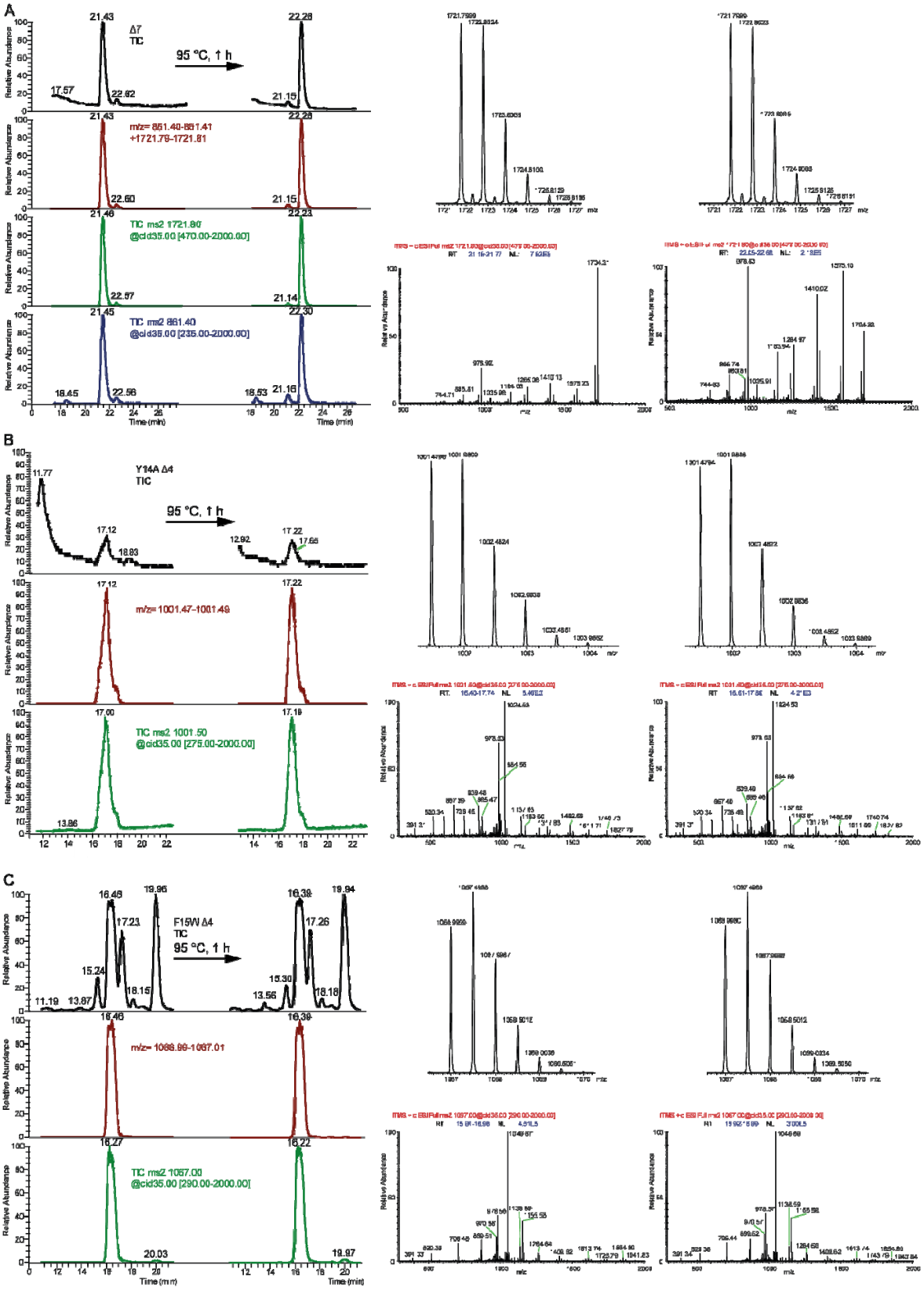
Heterologous expression of the WT and mutated astexin-1 biosynthesis cluster

Torsion angle constrains for NMR structure elucidation

References

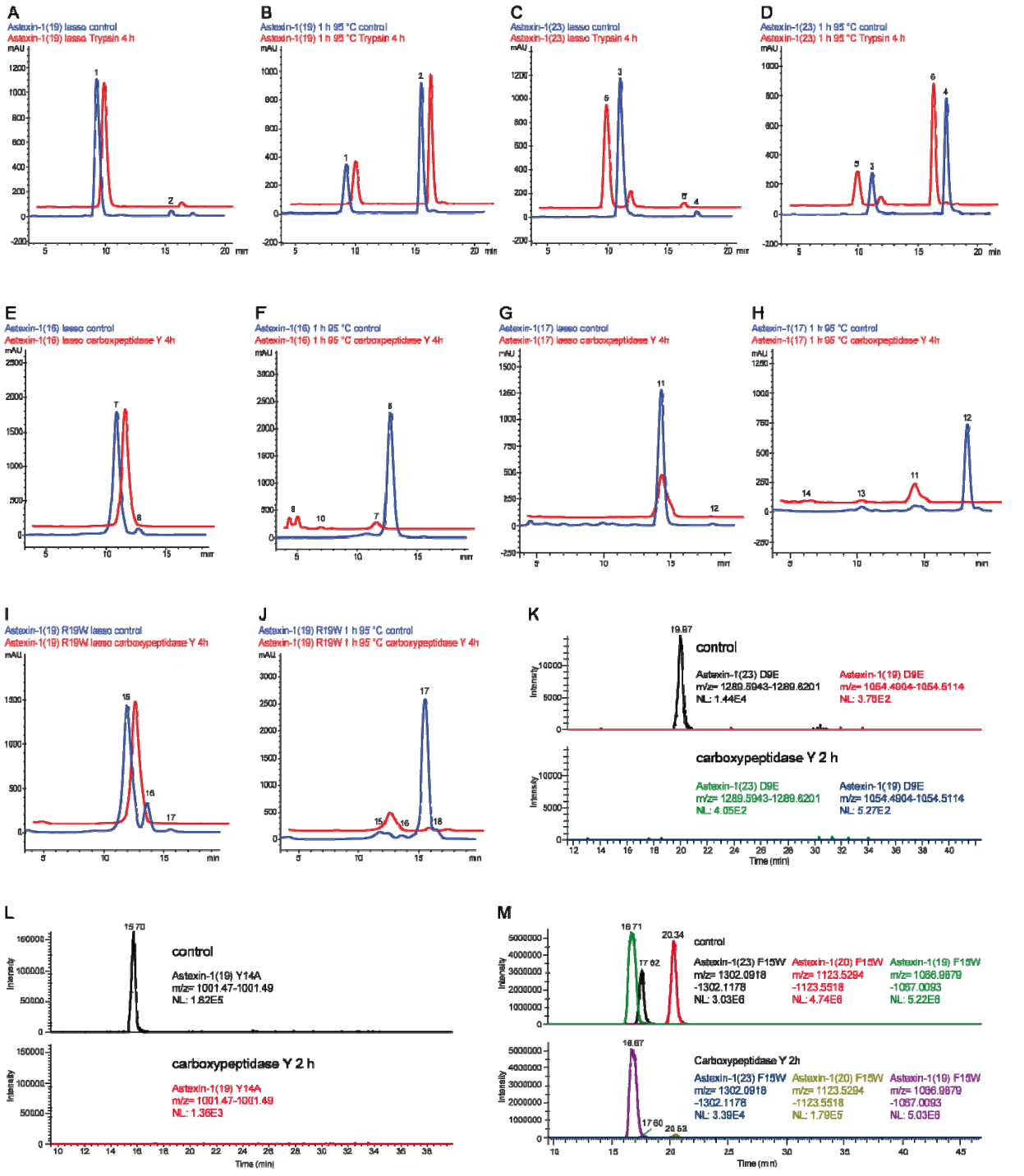


Supplementary Figure S 1, related to Figure 1. Purification of astexin-1 products. **A** Preparative HPLC chromatogram. **1** Astexin-1(18), (19), (21), (22) and (23), **2** Astexin-1(20). **B** Chromatogram of the first analytical HPLC. **3** Astexin-1((19), **4** Astexin-1(18) and (23), **5** Astexin-1(21) and (22), **6** Astexin-1(20). **C** Chromatogram of the second analytical HPLC. **7** Astexin-1((18), **8** Astexin-1(23)

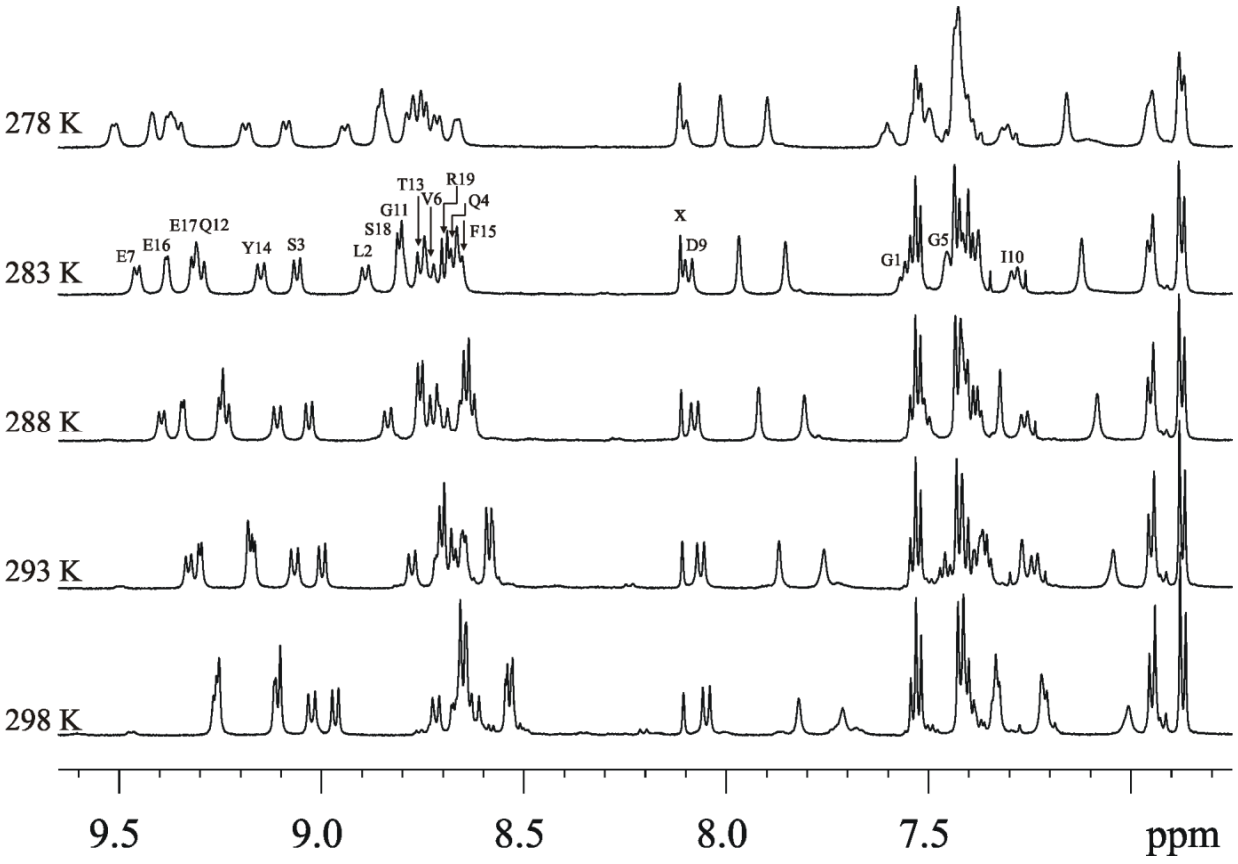


Supplementary Figure S 2, related to Figure 2. Exemplary heat stability of three different astexin-1 variants. **A** astexin-1(16) Δ7 variant as an example of a heat labile variant, **B** astexin-1(19) Y14A variant as an example of an already unfolded variant, **C** astexin-1(19) F15W variant as the only heat stable variant. Each sample is shown as TIC and EIC on the left, MS and MS² spectra in the middle

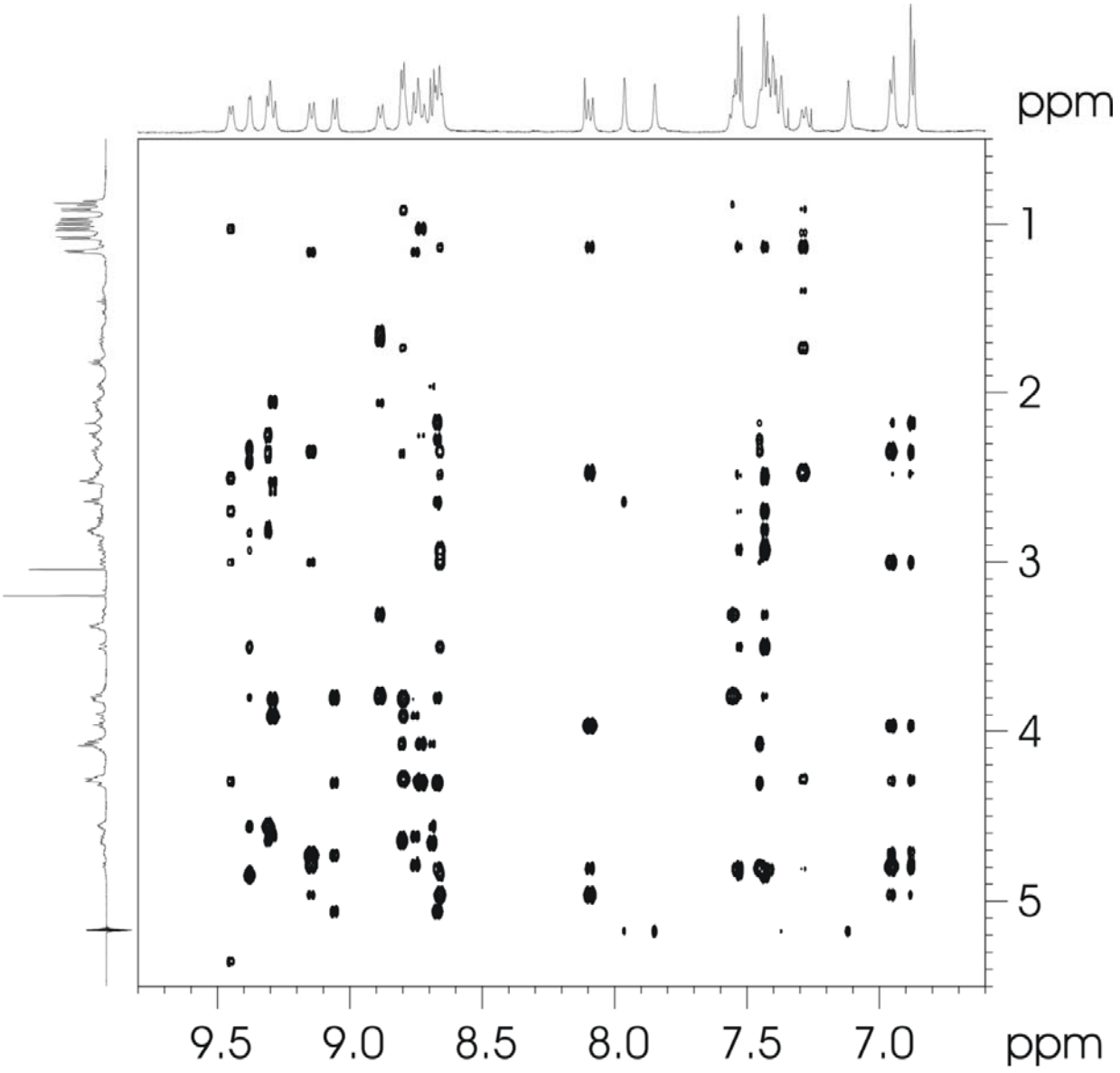
and on the right. MS and MS² spectra in the middle correspond to the peaks before, while those on the right correspond to the peaks after heat treatment.



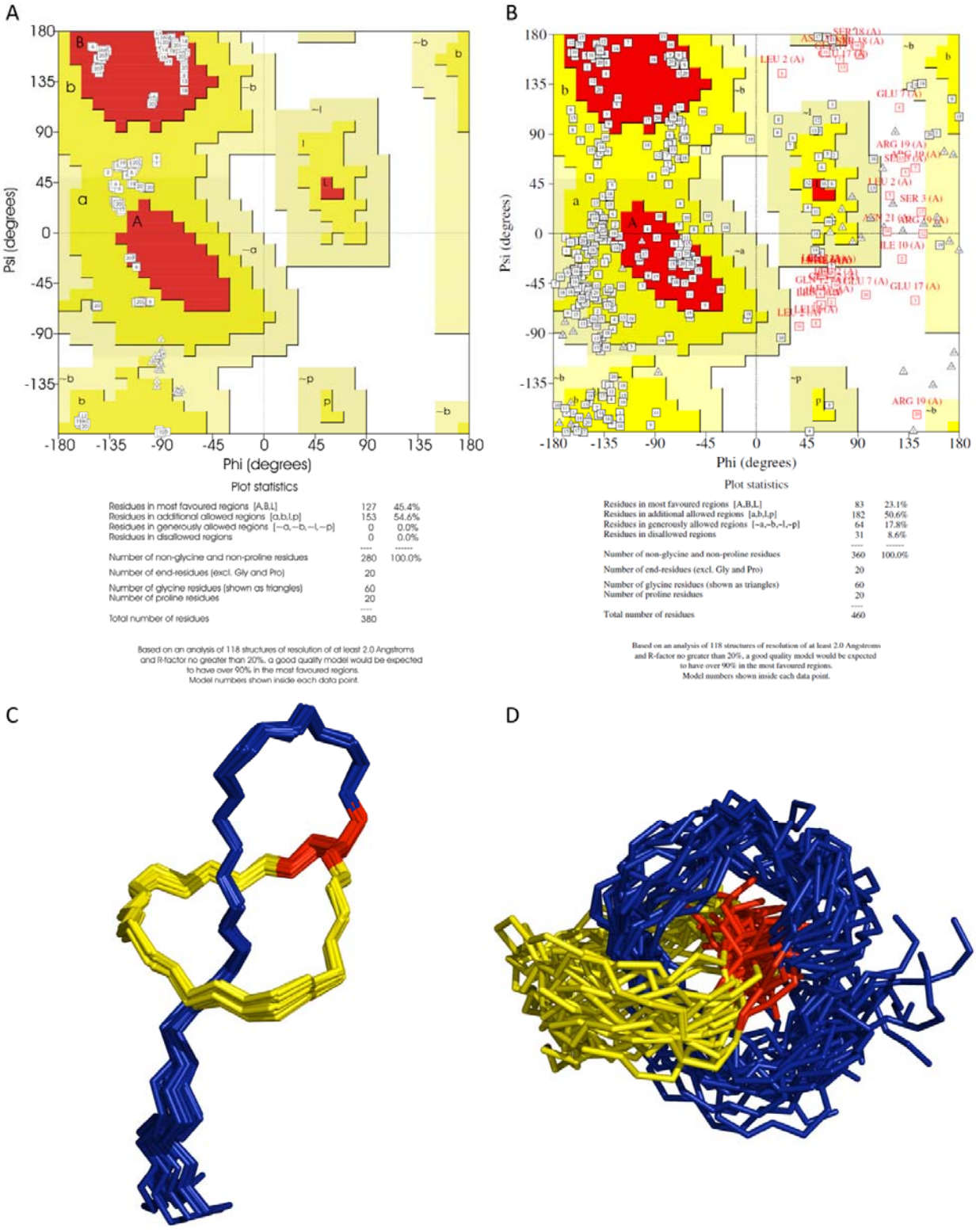
Supplementary Figure S 3, related to Figure 3. Stability of native astexin-1 and variants against proteolytic degradation shown as HPLC chromatograms (A – J) or EIC (K – M). Trypsin digestion of WT astexin-1(19) **A** before and **B** after heat treatment. **1** astexin-1(19) lasso, **2** astexin-1(19) unfolded. Trypsin digestion of WT astexin-1(23) **C** before and **D** after heat treatment. **3** astexin-1(23) lasso, **4** astexin-1(23) unfolded, **5** astexin-1(19) lasso, **6** astexin-1(19) unfolded. Carboxypeptidase Y digestion of the astexin-1 Δ7(16) variant **E** before and **F** after heat treatment. **7** astexin-1(16) lasso, **8** astexin-1(16) unfolded, **9** astexin-1(11) and (13), **10** astexin-1(10). Carboxypeptidase Y digestion of the astexin-1 Δ6(17) variant **G** before and **H** after heat treatment. **11** astexin-1(17) lasso, 1**2** astexin-1(17) unfolded, 1**3** astexin-1(14), 1**4** astexin-1(12) and (13). Carboxypeptidase Y digestion of the R19W astexin-1 variants (19) and (21) **I** before and **J** after heat treatment. **15** astexin-1(19), **16** astexin-1(21), **17** astexin-1(19) unfolded, **18** astexin-1(21) unfolded. Carboxypeptidase Y digestion of variants **K** D9E, **L** Y14A and **M** F15W shown as EIC.



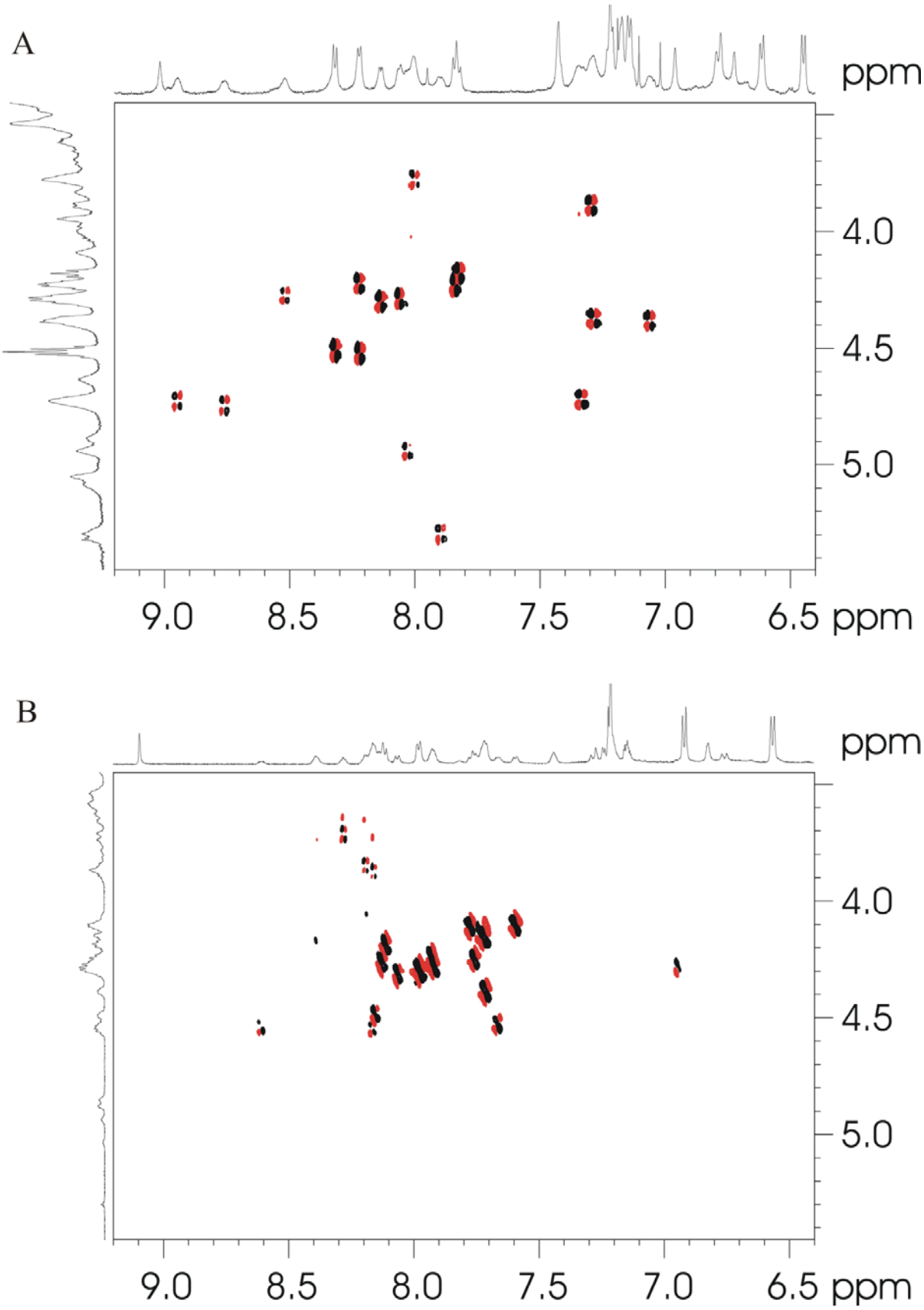
Supplementary Figure S 4, related to Figure 5. Variable temperature ^{1}H spectra in the region 9.65 – 6.75 ppm of astexin-1(19) in H2O/D2O (9:1). The labels are signal assignments of the amide protons, whereby x stands for an impurity.



Supplementary Figure S 5, related to Figure 6. Section of NOESY spectrum (mixing time 300 ms) of astexin-1(19) in H_2O/D_2O at 283 K.



Supplementary Figure S 6, related to Figure 7. Quality comparison of the here presented and the recently published structure (Maksimov, et al., 2012). **A** Ramachandran Plot of the 20 lowest energy structures of astexin-1(19) created by PROCHECK NMR. **B** Ramachandran plot of the 20 lowest energy structures of astexin-1(23) from Maksimov et al. 2012. **C** Superimposition of the 20 minimum energy structures of astexin-1(19) in the same color code as Figure 6A. **D** Superimposition of the 20 minimum energy structures of astexin-1(23) in the same color code created with PyMol from the PDB file 2LTI (Maksimov, et al., 2012).



Additional Figure 1 Section of DQF-COSY spectra of **A** lasso astexin-1(23) and **B** unfolded astexin-1(19) in DMSO-d_6 at 295 K

Supplementary Table S 1, related to Figure 4. Summary of the results of the mutational analysis of the astexin-1 system including production, heat stability and stability against carboxypeptidase Y.

mutant	rel. production by Integration of UV in %	thermostable	stability against carboxypeptidase Y digestion
WT	100 ± 23	no	unfolded unstable
T-2A	detected my MS	*	*
T-2V	not detected by MS	-	-
T-2S	detected by MS	*	*
R-1A	20 ± 6	*	*
G1C	not detected by MS	-	-
D9E	detected by MS	already unfolded	unstable
I10A	4.1 ± 0.2	no	-
G11A	19 ± 2	no	-
Q12A	72 ± 34	no	-
T13A	85 ± 3	no	-
Y14A	detected by MS	already unfolded	unstable
F15A	not detected by MS	-	-
E16A	14.0 ± 0.8	no	-
E17A	33 ± 5	no	-
S18A	44 ± 8	no	-
R19A	72 ± 8	no	-
I20A	2.5 ± 0.6	no	-
N21A	22.3 ± 1.6	no	-
Q22A	67 ± 9	no	-
D23A	10.4 ± 1.3	no	-
Δ1 aa	33 ± 3	no	-
Δ2 aa	28.8 ± 0.9	no	-
Δ3 aa	1.9 ± 0.2	no	-
Δ4 aa	0.7 ± 0.2	no	-
Δ5 aa	16.7 ± 0.6	no	-
Δ6 aa	18.1 ± 0.3	no	unfolded unstable
Δ7 aa	2.3 ± 0.2	no	unfolded unstable
Δ8 aa	detected by MS	already unfolded	-

mutant	rel. production by Integration of UV in %	thermostable	stability against carboxypeptidase Y digestion
Δ9 aa	not detected by MS	-	-
Y14W	2.7 ± 0.3	no	-
F15W	24 ± 5	yes	stable
R19W	18 ± 4	no	unfolded unstable

* = sequence equals the WT; - = was not tested

Supplementary Table S 2. Assignment of ^{1}H Signals (ppm) of astexin-1 (23) in H_2O/D_2O (9:1) at 288 K

AA	NH	αH	βH	others
Gly1	7.531	3.789; 3.307	/	/
Leu2	8.835	5.174	2.082; 1.684	γCH: 1.637
				δCH_3: 1.090; 0.980
Ser3	9.033	5.074	4.292; 3.811	
Gln4	8.670	4.809	2.283; 2.184	γCH_2: 2.642
				εNH_2: 7.920; 7.323
Gly5	7.434	4.327; 4.077	/	/
Val6	8.704	4.297	2.262	γCH_3: 1.036; 1.009
Glu7	9.386	5.355	2.694; 2.493	γCH_2: 2.889; 2.816
Pro8	/	3.974	2.478; 2.083	γCH_2: 2.396; 2.183
				δCH_2: 4.115; 4.697
Asp9	8.082	4.808	2.489; 1.146	/
Ile10	7.268	4.285	1.736	γCH_2: 1.408; 1.044
				γCH_3: 0.924; δCH3: 0.886
Gly11	8.762	3.930; 3.813	/	/
Gln12	9.233	4.626	2.060	γCH_2: 2.589; 2.527
				εNH_2: 7.809; 7.085
Thr13	8.728	4.794	4.730	γCH_3: 1.174
Tyr14	9.109	4.965	3.000; 2.352	δCH: 6.957; εCH: 6.884
Phe15	8.661	4.848	3.494; 2.930	δCH: 7.433; εCH: 7.536
Glu16	9.345	4.567	2.399; 2.339	γCH_2: 2.824
Glu17	9.246	4.627	2.352; 2.250	γCH_2: 2.779; 2.744
Ser18	8.762	4.641	4.062	/
Arg19	8.692	4.582	1.947	γCH_2: 2.069; 1.809;
				δCH_2: 3.383; εNH: 7.386
Ile20	8.409	4.328	2.029	γCH_2: 1.626; 1.364
				γCH_3: 1.077; δCH3: 1.046
Asn21	8.776	4.873	3.040; 2.930	δNH_2: 7.854; 7.134
Gln22	8.605	4.538	2.331; 2.156	γCH_2: 2.544
				εNH_2: 7.766; 7.105
Asp23	8.620	4.867	3.135; 3.084	/

Supplementary Table S 3. Assignment of ^{1}H Signals (ppm) of astexin-1 (19) in H_2O/D_2O (9:1) at 283 K

AA	NH	αH	βH	others
Gly1	7.555	3.786; 3.308	/	/
Leu2	8.885	5.177	2.059; 1.683	γCH: 1.637
				δCH_3: 1.083; 0.970
Ser3	9.059	5.060	4.307; 3.801	/
Gln4	8.670	4.804	2.270; 2.174	γCH_2: 2.641
				εNH_2: 7.963; 7.370
Gly5	7.452	4.307; 4.074	/	/
Val6	8.731	4.284	2.254	γCH_3: 1.031; 0.999
Glu7	9.449	5.352	2.698; 2.501	γCH_2: 2.892; 2.807
Pro8	/	3.967	2.476; 2.081	γCH_2: 2.389; 2.179
				δCH_2: 4.109; 4.705
Asp9	8.094	4.803	2.470; 1.132	/
Ile10	7.286	4.281	1.729	γCH_2: 1.388; 1.045
				γCH_3: 0.924; δCH3: 0.877
Gly11	8.799	3.907; 3.804	/	/
Gln12	9.290	4.618	2.053	γCH_2: 2.593; 2.523
				εNH_2: 7.849; 7.117
Thr13	8.753	4.800	4.752	γCH_3: 1.161
Tyr14	9.147	4.959	3.000; 2.345	δCH: 6.953; εCH: 6.874
Phe15	8.660	4.843	3.500; 2.930	δCH: 7.430; εCH: 7.527
Glu16	9.379	4.561	2.407; 2.330	γCH_2: 2.828
Glu17	9.308	4.64	2.360; 2.248	γCH_2: 2.797; 2.751
Ser18	8.803	4.659	4.077	/
Arg19	8.689	4.556	1.961	γCH_2: 2.113; 1.822;
				δCH_2: 3.380; εNH: 7.400

Supplementary Table S 4. Structural statistics for the family of 20 structures selected to represent the structure of astexin-1(19) in H_2O/D_2O (9:1)

Restraining Constraints	Constraints Violations
Total: 202	Distance violations, >0.5 Å: 0
distance, i=j: 105	RMS deviations: 0.02 Å
distance, |i-j|=1: 40	Dihedral violations, > 5°: 0
distance, |i-j|>1: 24	RMS deviation: 1.8°
dihedral: 33	Average pairwise RMS deviation (Leu^2-Ser^{18})
hydrogen bond: 0	Backbone atoms: 0.03 Å
Constraints/residue: 10.6	All heavy atoms: 0.30 Å

Supplementary Table S 5. M9 Vitamin Mix

component	amount
biotin	0.2 g
choline chloride	1.0 g
disodium adenosine 5'-triphosphate	0.3 g
folic acid	1.0 g
myo-inositol	2.0 g
nicotinamide	1.0 g
panthothenic acid	1.0 g
pyridoxal hydrochloride	1.0 g
riboflavin	0.1 g
thiamine	1.0 g
H_2O	ad 300 mL

10 M NaOH were added to the resulting suspension until everything was dissolved.
The solution was then sterile filtered and stored at 4°C

Supplementary Table S 6. Oligonucleotide Primers

name	sequence
Astex0830A_NdeI_FP	GGATTAC CAT ATG CAT ACC CCC ATT ATT TCG GAA ACG GTC
Astex0832C_XhoI_RP	ATAT CTC GAG TTA GGT CCG CGG ACG GCG C
Astex3626A1_NdeI_FP	GGATTAC CAT ATG ACC AAG CGT ACG ACC ATC GCC
Astex3623C_XhoI_RP	ATAT CTC GAG CTA GGG TGG CCC CGC CCG AG
AE08_rth_tail_all_FP	TGA CGA CGG GAA GGC GCG CAC AC
AE08_-9aa_RP	GTA GGT CTG GCC GAT ATC GGG CTC GAC AC
AE08_-8aa_RP	GAA GTA GGT CTG GCC GAT ATC GGG CTC G
AE08_-7aa_RP	TTC GAA GTA GGT CTG GCC GAT ATC GGG C
AE08_-6aa_RP	CTC TTC GAA GTA GGT CTG GCC GAT ATC GGG
AE08_-5aa_RP	AGA CTC TTC GAA GTA GGT CTG GCC GAT ATC G
AE08_-4aa_RP	GCG AGA CTC TTC GAA GTA GGT CTG GCC G
AE08_-3aa_RP	GAT GCG AGA CTC TTC GAA GTA GGT CTG GCC
AE08_-2aa_RP	GTT GAT GCG AGA CTC TTC GAA GTA GGT CTG G
AE08_-1aa_RP	CTG GTT GAT GCG AGA CTC TTC GAA GTA GGT C
AE08_R19A_RP	GTC CTG GTT GAT G**GC** AGA CTC TTC GAA GTA GGT CTG GCC
AE08_F15A_RP	GTC CTG GTT GAT GCG AGA CTC TTC G**GC** GTA GGT CTG GCC
AE08_Y14A_RP	GTC CTG GTT GAT GCG AGA CTC TTC GAA G**GC** GGT CTG GCC
AE08_R19W_RP	GTC CTG GTT GAT **C**C**A** AGA CTC TTC GAA GTA GGT CTG GCC
AE08_F15W_RP	GTC CTG GTT GAT GCG AGA CTC TTC **C**C**A** GTA GGT CTG GCC
AE08_ring_all_FP	GGC CAG ACC TAC TTC GAA GAG TCT CGC ATC
AE08_G1C_RP	GAT ATC GGG CTC GAC ACC CTG GCT GAG AC**A** GCG GGT
AE08_D9E_RP	GAT **C**TC GGG CTC GAC ACC CTG GCT GAG ACC GCG GGT
AE08_Leader_all_FP	GGT CTC AGC CAG GGT GTC GAG CCC GAT ATC
Astex0830_R-1A_RP	G**GC** GGT CTC AGC TGA AGC CTT GCC CAG AAC

name	sequence
Astex0830_T-2A_RP	GCG GG**C** CTC AGC TGA AGC CTT GCC CAG AAC
AE08_T-2S_RP	GCG G**C**T CTC AGC TGA AGC CTT GCC CAG AAC
AE08_T-2V_RP	GCG G**AC** CTC AGC TGA AGC CTT GCC CAG AAC
AE08SLIMTail1FP	GAG TCT CGC ATC AAC CAG GAC TGA CGA CG
AE08SLIMTail1RP	ATC GGG CTC GAC ACC CTG GCT GAG ACC
AE08SLIM_I10A_FP	**GC**C GGC CAG ACC TAC TTC GAA GAG TCT CGC ATC AAC CAG GAC TGA CGA CG
AE08SLIM_I10A_RP	TTC GAA GTA GGT CTG GCC G**GC** ATC GGG CTC GAC ACC CTG GCT GAG ACC
AE08SLIM_G11A_FP	ATC G**C**C CAG ACC TAC TTC GAA GAG TCT CGC ATC AAC CAG GAC TGA CGA CG
AE08SLIM_G11A_RP	TTC GAA GTA GGT CTG G**G**C GAT ATC GGG CTC GAC ACC CTG GCT GAG ACC
AE08SLIM_Q12A_FP	ATC GGC **GC**G ACC TAC TTC GAA GAG TCT CGC ATC AAC CAG GAC TGA CGA CG
AE08SLIM_Q12A_RP	TTC GAA GTA GGT C**GC** GCC GAT ATC GGG CTC GAC ACC CTG GCT GAG ACC
AE08SLIM_T13A_FP	ATC GGC CAG **G**CC TAC TTC GAA GAG TCT CGC ATC AAC CAG GAC TGA CGA CG
AE08SLIM_T13A_RP	TTC GAA GTA GG**C** CTG GCC GAT ATC GGG CTC GAC ACC CTG GCT GAG ACC
AE08SLIM_Y14W_FP	ATC GGC CAG ACC T**GG** TTC GAA GAG TCT CGC ATC AAC CAG GAC TGA CGA CG
AE08SLIM_Y14W_RP	TTC GAA **CC**A GGT CTG GCC GAT ATC GGG CTC GAC ACC CTG GCT GAG ACC
AE08SLIM_E16A_FP	ATC GGC CAG ACC TAC TTC G**CC** GAG TCT CGC ATC AAC CAG GAC TGA CGA CG
AE08SLIM_E16A_RP	**GG**C GAA GTA GGT CTG GCC GAT ATC GGG CTC GAC ACC CTG GCT GAG ACC
AE08SLIMTail2FP	TGA CGA CGG GAA GGC GCG CAC ACT G
AE08SLIMTail2RP	TTC GAA GTA GGT CTG GCC GAT ATC GGG CTC
AE08SLIM_E17A_FP	G**C**G TCT CGC ATC AAC CAG GAC TGA CGA CGG GAA GGC GCG CAC ACT G

name	sequence
AE08SLIM_E17A_RP	GTC CTG GTT GAT GCG AGA C**G**C TTC GAA GTA GGT CTG GCC GAT ATC GGG CTC
AE08SLIM_S18A_FP	GAG **G**CT CGC ATC AAC CAG GAC TGA CGA CGG GAA GGC GCG CAC ACT G
AE08SLIM_S18A_RP	GTC CTG GTT GAT GCG AG**C** CTC TTC GAA GTA GGT CTG GCC GAT ATC GGG CTC
AE08SLIM_I20A_FP	GAG TCT CGC **GC**C AAC CAG GAC TGA CGA CGG GAA GGC GCG CAC ACT G
AE08SLIM_I20A_RP	GTC CTG GTT G**GC** GCG AGA CTC TTC GAA GTA GGT CTG GCC GAT ATC GGG CTC
AE08SLIM_N21A_FP	GAG TCT CGC ATC **GC**C CAG GAC TGA CGA CGG GAA GGC GCG CAC ACT G
AE08SLIM_N21A_RP	GTC CTG G**GC** GAT GCG AGA CTC TTC GAA GTA GGT CTG GCC GAT ATC GGG CTC
AE08SLIM_Q22A_FP	GAG TCT CGC ATC AAC **GC**G GAC TGA CGA CGG GAA GGC GCG CAC ACT G
AE08SLIM_Q22A_RP	GTC C**GC** GTT GAT GCG AGA CTC TTC GAA GTA GGT CTG GCC GAT ATC GGG CTC
AE08SLIM_D23A_FP	GAG TCT CGC ATC AAC CAG G**C**C TGA CGA CGG GAA GGC GCG CAC ACT G
AE08SLIM_D23A_RP	G**G**C CTG GTT GAT GCG AGA CTC TTC GAA GTA GGT CTG GCC GAT ATC GGG CTC

Introduced restriction sites are underlined; introduced mutations are shown in bold.

Supplementary Methods

Strains and general methods. *Asticcacaulis excentricus* CB48 (DSM 4724) was purchased from the German Collection for Microorganisms and Cell Cultures (DMSZ). *E. coli* TOP10, which was used for cloning, and *E. coli* BL21 (DE3), which was used for heterologous expression, were purchased from Invitrogen. Oligonucleotides and carboxypeptidase Y were purchased from Sigma Aldrich. Trypsin was purchased from Promega. Restriction enzymes, Phusion polymerase and T4 DNA Ligase were purchased from New England Biolabs. DNA Sanger Sequencing was performed by GATC Biotech AG (Konstanz) to determine the identity of constructed plasmids and mutants.

Cloning of the astexin-1 biosynthesis cluster. Genomic DNA of *Asticcacaulis excentricus* CB48 was used to amplify the *atxABC* gene cluster with PCR with the first two primers of Supplementary Table S6. PCR was performed with Phusion Polymerase following the instructions of the manufacturer. PCR was subjected to gel electrophoresis and amplicons with the correct size were excised, purified and subsequently digested with NdeI and XhoI. pET41a vector was digested in the same way and the fragments were ligated with T4 DNA Ligase. Transformation was performed with TOP10 cells for cloning and BL21(DE3) for expression.

Construction of astexin-1 mutants. The a*stex222xABC*pET41a plasmid was used as template for all performed mutations. Mutagenesis was either done with a modified protocol of the site directed mutagenesis with inverse PCR (Hemsley, et al., 1989) or with the site-directed-ligation-independent-mutagenesis (SLIM) protocol (Chiu, et al., 2008). Primers for each mutant or mutant group are shown in Supplementary Table S6.

Heterologous expression of the WT and mutated astexin-1 biosynthesis cluster. *E. coli* BL21(DE3) cells were transformed with the plasmid *astex222xABC*pET41 or mutants thereof and

grown overnight at 37°C in LB medium containing kanamycin (50 μg/mL). M9 minimal medium supplemented with M9 vitamin mix (Supplementary Table S5) and containing kanamycin (50 μg/mL) was inoculated with the overnight culture to an OD_{600} of 0.01. The culture was grown at 37 °C to an OD_{600} of 0.6 and the expression was induced with the addition of IPTG to a final concentration of 0.1 mM. Expression was continued overnight and the cells were harvested by centrifugation. Pellets were extracted as stated in the purification section and extracts were subsequently applied to HPLC for purification or analyzed by a HPLC-MS system.

Torsion angle constraints for NMR structure elucidation. Only those unambiguous coupling constants were used. Thus, $^{3}J_{HN\alpha} \geq 9$ Hz was observed for Leu2, Ser3, Val6, Glu7, Asp9, Ile10, and Glu12 – Phe15. The torsion angles ϕ of these residues were restrained to -120° ± 30°. For the residues Gly1, Gln4, Gly5, Gly11, and Glu16 – Arg19, $^{3}J_{HN\alpha} < 9$ Hz was detected. Thus their torsion angles ϕ were restrained to -70° ± 30°. Stereospecific assignment of the following prochiral β-methylene protons were fulfilled by measuring $^{3}J_{\alpha\beta}$ and analyzing patterns of the intraresidual NOE interactions $d_{\alpha\beta}$ and $d_{N\beta}$ (Wagner, 1990): Leu2, Ser3, Asp9, Try14, and Phe15 in (t^2g^3); Gln4, Glu7, and Glu17 in (g^2g^3). For $t^2g^3$and g^2g^3 conformations around the Cα-Cβ bond the torsion angle χ^1 was constrained in the range of -60 ± 30°and 60 ± 30° respectively.

References

Chiu, J., Tillett, D., Dawes, I.W., and March, P.E. (2008). Site-directed, Ligase-Independent Mutagenesis (SLIM) for highly efficient mutagenesis of plasmids greater than 8kb. J. Microbiol. Methods 73, 195-198.

Hemsley, A., Arnheim, N., Toney, M.D., Cortopassi, G., and Galas, D.J. (1989). A simple method for site-directed mutagenesis using the polymerase chain reaction. Nucleic Acids Res. 17, 6545-6551.

Maksimov, M.O., Pelczer, I., and Link, A.J. (2012). Precursor-centric genome-mining approach for lasso peptide discovery. Proc. Natl. Acad. Sci. USA 109, 15223-15228.

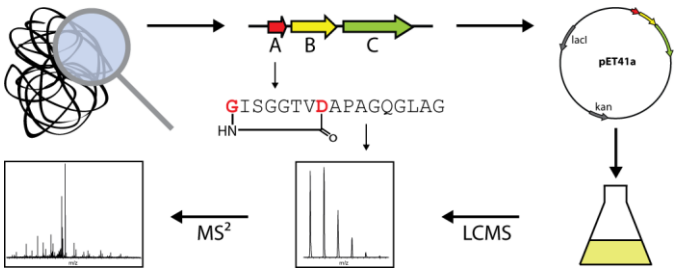
Wagner, G. (1990). Nmr Investigations of Protein-Structure. Progress in Nuclear Magnetic Resonance Spectroscopy 22, 101-139.

Lasso Peptides from Proteobacteria: Genome Mining Employing Heterologous Expression and Mass Spectrometry

*Julian D. Hegemann, Marcel Zimmermann, Shaozhou Zhu, Dennis Klug, Mohamed A. Marahiel**

Department of Chemistry, Biochemistry, Philipps-University Marburg, Hans-Meerwein-Strasse 4 and LOEWE-Centre for Synthetic Microbiology, D-35032, Marburg, Germany

*marahiel@staff.uni-marburg.de



Supporting Information

Supplementary Table S1. Primers for the cloning of the investigated gene clusters. Introduced restriction sites are underlined.

name	sequence
BurhA_NdeI_FP	GGATTAC CAT ATG AAC AAG CAA CAA CAA GAG TCG GGC CTG C
BurhC_SpeI_RP	ATAT ACT AGT TAT TGC CAA CTT TGA ATA AAC AGC TCG GCA GC
CK31_A1_NdeI_FP	GGATTAC CAT ATG GAA CGG ATC GAA GAC CAC ATC GAC GAC GAA CTG ATC GAC CTG GGC G
CK31_C_EcoRI_RP	ATAT GAA TTC TCA GAC ATC CTC CCA GGA TCG CAC CCA GGA CTC GAC GTC GGC C
PzucA_NdeI_FP	GGATTAC CAT ATG ACC AGG CTT CTC AAT CTG ATG AGC GTT CGG CTG
PzucC_HindIII_RP	ATAT AAG CTT TCA GGC CGA CCA CCG CCG GAC CCA GGC
RhotA_NdeI_FP	GGGAATTC CAT ATG ACG CAA TCG CAG GAG ACT GAG ATG GAC ACG AAC GAG AAC
RhotC_HindIII_RP	ATAT AAG CTT TTA GGG GTA TTG GGA CTG CTG CCG TAC CCA GTT CTC CAA GGC
RugeA_NdeI_FP	GGGAATTC CAT ATG AAG GAG TTC GCC ATG GAC GAA GAG CTG GAA CTG GAG ATC GTC GAC C
RugeC_HindIII_RP	ATAT AAG CTT TCA AGC GAG CGG ACG AGG CGC GGC CGC G
Sjap1_A_NdeI_FP	GGGAATTC CAT ATG GAA CGT GAC AAT GAC GTG ATC GAA CTC GGC GCG GTC AGC GTG
Sjap1_C_XhoI_RP	ATAT CTC GAG CTA ATC CTG GGT CTG CCC CGC CAC CCA TGC CTC GAC ATC GGC
Sjap2_A_NdeI_FP	GGGAATTC CAT ATG GAT CGT CAT GAC AAC TCC GAG GTC GAT GAG ATC ATC GAC CTC GGC ACC
Sjap2_C_HindIII_RP	ATAT AAG CTT CTA GGC AGG TTG CGA CCT GGC CCA CGC CTC GAC GTC G
SyanA_NdeI_FP	GGGAATTC CAT ATG GAA CGC AAC AGC GAA GAC CGC CGT GAC GAC GTC GTC G
SyanC_HindIII_RP	ATAT AAG CTT TCA GCC GCG ACG GGC AGC GAG CCA GGC TTC GAC
Sala1_A_NdeI_FP	GGAATTC CAT ATG AAA GAC TTC AAC GAA CTG ATC GAC C
Sala1_C_XhoI_RP	CCG CTC GAG TCA ATT TCC CTC CCA ATA TCG GGC C
Sala2_A_NdeI_FP	GGAATTC CAT ATG GAA CGC ACC GAA GTA ATC GAG G
Sala2_C_EcoRI_RP	GGC GAA TTC TTA TCG GCC CAA TGC GCT ATC G

Supplementary Table S2. Primers for the introduction of the *E. coli* optimized RBS in between the precursor genes and the genes encoding the processing enzymes. All primers were 5'-phosphorylated and the RBS sequence is underlined.

name	sequence
Burh_RBS_FP	AGAGGAGAAATTAACC ATG CAG CAT GGC GCT TGG AGA CAA AGT CAG CTA TG
Burh_RBS_RP	TTA TTC GTC GTC GGA GTC AAT GCG ATC ACT CCA GCG
CK31_A1A2A3_RBS_FP	AGAGGAGAAATTAACC ATG CGC GTA GCC GTT CCC GAT CAT CTC GCA TAC
CK31_A1A2A3_RBS_RP	CTA GTC TTC CGA CAG GCC CAT GGG CGG GCG
CK31_A1_RBS_FP[a]	AGAGGAGAAATTAACC ATG CGC GTA GCC GTT CCC GAT CAT CTC GCA TAC
CK31_A1_RBS_RP[a]	TTA GTC CCG GGA CAG GCC CGT GGG CTC CCG G
CK31_A2_RBS_FP[a]	ATG CAA CGG ATC ATC GAC GAG ACC ACG GAC GGT CTG
CK31_A2_RBS_RP[a]	ATG TAT ATC TCC TTC TTA AAG TTA AAC AAA ATT ATT TCT AGA GGG GAA TTG
CK31_A3_RBS_FP[a]	ATG GAA TTT GAA GGC ATC CCT TCC CCG GAC GCG CGG ATC GAT CTC
CK31_A3_RBS_RP[a]	ATG TAT ATC TCC TTC TTA AAG TTA AAC AAA ATT ATT TCT AGA GGG G
Pzuc_RBS_FP	AGAGGAGAAATTAACC ATG CCC GCG CTG CAC CCG AAA TCC TTT CTC GTC
Pzuc_RBS_RP	TCA GAC GTC GAA CGG CTT GTT CAG GTC TTC GAA GTC ACC
Rhot_RBS_FP	AGAGGAGAAATTAACC ATG ACC AGC TAT TGG CTA CAC GAC GAC CTC TCA TTC TGC C
Rhot_RBS_RP	TTA CTC GGT GAT TCC CGG AGT GGC GGC GTG GCC
Ruge_RBS_FP	AGAGGAGAAATTAACC ATG TCG GCA ACG ACA CGC CGT CTA CAC CCG AGG CCG ATC TTG
Ruge_RBS_RP	TCA GAC GCG CTG CGG GAA GGC CGG GTT GTC TTC CGA G
Sjap1_RBS_FP	AGAGGAGAAATTAACC ATG ACC GGC TTT ACA CTG CGC GAT GGT CTG AGC
Sjap1_RBS_RP	TCA GTC GTC GGA AAG ACC GAA ATT ATT TTC ACC AAG CCC GAC ATC
Sjap2_RBS_FP	AGAGGAGAAATTAACC ATG GCG CTG GCA CTG CGT GAC CAG CTC TC
Sjap2_RBS_RP	TCA GTC GTC AGA AAT GCC GCC GAT CAG GTT TTT GGG C
SyanA_RBS_FP	AGAGGAGAAATTAACC ATG GCC CGC GCG GAC ATA AAT TGG TGC GAT GTC GGA GGG C
SyanC_RBS_RP	TCA GTC GTC GAG GAT ACC GGC AAG GCC CTG GCC GGC AG
Sala1_RBS_FP	AGAGGAGAAATTAACC ATG TCC GCC ATG CTG CGA TCG GGC CTG CAC
Sala1_RBS_RP	TCA GTC GTC GGC CGT GAT GCC GCC GGC
Sala2_RBS_FP	AGAGGAGAAATTAACC ATG CCT GCA ATG CAC TAT AGG ACG GCA CCC GGC
Sala2_RBS_RP	TCA GTC CTG CGC GAT TCC GGT GAG GAA CTG C

[a]*CK31_A1A2A3BC* pET41a was used as template

Supplementary Table S3. M9 Vitamin Mix.

component	amount
choline chloride	1.0 g
folic acid	1.0 g
pantothenic acid	1.0 g
nicotinamide	1.0 g
myo-inositol	2.0 g
pyridoxal hydrochloride	1.0 g
thiamine	1.0 g
riboflavin	0.1 g
disodium adenosine 5'-triphosphate	0.3 g
biotin	0.2 g
	ad 300 mL ddH_2O[a]

[a]10 M NaOH was slowly added to the resulting suspension until everything was dissolved (pH ~12). Afterwards the solution was sterile filtered and stored at 4 °C until needed.

Supplementary Table S4a. Genome mining results I. Overview over identified gene clusters.

strain	cluster organization	GntR present?	TonB present?	GenBank accession no. of the B/B2 protein
Asticaccaulis excentricus CB48	ABC	yes	yes	YP_004088036.1
Asticaccaulis excentricus CB48	AABC	yes	yes	ADU14097.1
Breviundimonas diminuta ATCC 11568	ABC	yes	yes	ZP_08267984.1
Burkholderia gladioli BSR3	ABCD	no	no	YP_004361612.1
Burkholderia mallei ATCC23344	ABC	no	no	YP_102882.1
Burkholderia oklahomensis C6786	ABCCD	no	no	ZP_02362859.1
Burkholderia oklahomensis EO147	ABCCD	no	no	ZP_02355720.1
Burkholderia pseudomallei 1106A	ABCD	no	no	YP_001066181.1
Burkholderia pseudomallei 1710b	ABCD	no	no	YP_333463.1
Burkholderia pseudomallei 668	ABCD	no	no	YP_001058938.1
Burkholderia pseudomallei 7894	ABCD	no	no	ZP_02481751.1
Burkholderia pseudomallei DM98	ABCD	no	no	ZP_02402883.1
Burkholderia pseudomallei K96243	ABCD	no	no	YP_108394.1
Burkholderia rhizoxinica HKI454	ABCD	no	no	YP_004027714.1
Burkholderia thailandensis E264	ABCD	no	no	YP_442959.1
Candidatus Odysella thessalonicensis L13	AABCD	no	no	ZP_08778228.1
Candidatus Odysella thessalonicensis L13	ABCD	no	no	ZP_08778197.1
Caulobacter crescentus CB15	ABC	no	yes	NP_421509.1
Caulobacter segnis ATCC 21756	AAABC	yes	yes	YP_003593637.1
Caulobacter sp. AP07	AABC	yes	yes	ZP_10749588.1
Caulobacter sp. K31	AABC	yes	yes	YP_001683611.1
Caulobacter sp. K31	AABC	yes	yes	YP_001683865.1
Caulobacter sp. K31	AAABC	yes	yes	YP_001676626.1
Chitinophaga pinensis DSM 2588	XB1KB2CAA	no	no	YP_003124540.1
Citromicrobium bathyomarinum JL354	ABC	no	no	ZP_06863107.1
Citromicrobium sp. JLT1363	CA*B2SKB1DD	no	no	ZP_08702804.1
Citromicrobium sp. JLT1363	B2XXB1KSCX	no	no	ZP_08702964.1
Escherichia coli AY25	ABCD	no	no	AAD28495.1
Escherichia coli MS 196-1	BC	no	no	ZP_07192991.1
Erythrobacter sp. NAP1	DBKACBX	no	no	ZP_01040581.1
Francisella cf. novicida 3523	XDB1B2CXX	no	no	YP_005823896.1
Francisella cf. novicida Fx1	XDB1B2CXXX	no	no	YP_005825876.1
Francisella novicida U112	XDB1B2CXX	no	no	YP_898576.1
Gallionella capsiferriformans ES-2	XB2XCXXB1XD	no	no	YP_003846630.1
Geobacter uraniireducens Rf4	A*CKKSB1B2D	no	no	YP_001230130.1
Hahella chejuensis KCTC 2396	BC	no	no	YP_432495.1
Haliangium ochraceum DSM 14365	ABC	no	no	YP_003265873.1
Legionella pneumophila 2300/99 Alcoy	A*XCB1B2DK	no	no	YP_003620325.1

Supplementary Table S4b. Genome mining results I. Overview over identified gene clusters.

strain	cluster organization	GntR present?	TonB present?	GenBank accession no. of the B/B2 protein
Legionella pneumophila str. Corby	A*XCB1B2DDK	no	no	YP_001252493.1
Legionella pneomophila subsp. *pneumophila* ATCC 43290	XCB1B2DK	no	no	YP_005187270.1
Legionella pneumophila subsp. *pneumophila*	XCB1B2DK	no	no	YP_006507101.1
Magnetospirillum magneticum AMB-1	A*BCD	no	no	YP_420827.1
Marine gamma proteobacterium HTCC2148	KXXXCADBX	no	no	ZP_05094211.1
Marinomonas mediterranea MMB-1	ACKSXBDD	no	no	YP_004312360.1
Mesorhizobium amorphea CCNWGS0123	AB1B2CKD	no	no	ZP_09090636.1
Mesorhizobium australicum WSM2073	AB1B2CKD	no	no	EHB69037.1
Mesorhizobium ciceri biovar biserrulae WSM1271	AB1B2CKD	no	no	YP_004139467.1
Mesorhizobium loti MAFF303099	AB1B2CKD	no	no	NP_105951.1
Mesorhizobium opportunistum WSM2075	AB1B2CKD	no	no	YP_004608836.1
Novosphingobium aromaticivorans DSM 12444	KB1B2CDX	no	no	YP_498511.1
Novosphingobium sp. AP12	ABC	no	yes	ZP_10744725.1
Novosphingobium sp. PP1Y	ABC	no	yes	YP_004532633.1
Novosphingobium sp. PP1Y	ABC	yes	yes	YP_004534330.1
Novosphingobium sp. Rr 2-17	ABC	yes	yes	ZP_10362270.1
Oligotropha carboxidovorans OM5	ACBD	no	no	YP_002290457.1
Phenylobacterium zucineum HLK1	ABC	yes	yes	YP_002131237.1
Phenylobacterium zucineum HLK1	ABC	yes	yes	YP_002131016.1
Photorhabdus asymbiotica ATCC 43949	BCD	no	no	YP_003041097.1
Pseudomonas geniculata N1	ABC	yes	yes	ZP_11209434.1
Psychromonas ingrahamii 37	B2DXB1XXC	no	no	YP_943976.1
Rhodanobacter sp. 115	ABC	yes	yes	ZP_10191822.1
Rhodanobacter sp. 116-2	ABC	yes	yes	ZP_10201279.1
Rhodanobacter sp. 2APBS1	ABC	yes	yes	ZP_08950893.1
Rhodanobacter thiooxydans LCS2	ABC	yes	yes	ZP_10205116.1
Rhodobacter sphaeroides 2.4.1	AB1B2CKD	no	no	YP_352615.1
Rhodobacter sphaeroides ATCC 17025	AB1B2CKD	no	no	YP_001168162.1
Rhodobacter sphaeroides ATCC 17029	AB1B2CKD	no	no	YP_001043097.1
Rhodobacter sphaeroides KD131	AB1B2CKD	no	no	YP_002525230.1
Rhodospirillum rubrum ATCC 11170	XB2XB1KSC	no	no	YP_428105.1
Rhodospirillum rubrum ATCC 11170	B1B2XC	no	no	YP_425910.1
Roseobacter denitrificans OCh 114	AB1B2CK	no	no	YP_683998.1
Roseobacter litoralis Och 149	AB1B2CK	no	no	YP_004689588.1
Rubrivivax benzoalyticus JA2	ABC	no	yes	ZP_08403836.1
Rubrivivax gelatinosus IL44	ABC	no	yes	YP_005435542.1
Shewanella sp. HN-41	B2XXDAB1KSC	no	no	ZP_08567205.1

Supplementary Table S4c. Genome mining results I. Overview over identified gene clusters.

strain	cluster organization	GntR present?	TonB present?	GenBank accession no. of the B/B2 protein
Sphingobium chlorophenolicum L-1	ABC	yes	yes	YP_004554865.1
Sphingobium japonicum UT26	ABC	yes	yes	YP_003546448.1
Sphingobium japonicum UT26	ABC	yes	yes	YP_003547070.1
Sphingobium sp. AP49	ABC	yes	yes	ZP_10544572.1
Sphingobium sp. SYK-6	ABC	no	yes	YP_004835823.1
Sphingobium yanoikuyae ATCC 51230	AB1B2CKD	no	no	EKU73221.1
Sphingobium yanoikuyae ATCC 51230	ABC	yes	yes	EKU73624.1
Sphingobium yanoikuyae XLDN2-5	AB1B2CKD	no	no	ZP_09906630.1
Sphingobium yanoikuyae XLDN2-5	ABC	yes	yes	ZP_09907189.1
Sphingobium yanoikuyae XLDN2-5	ABC	yes	yes	ZP_09906533.1
Sphingomonas sp. LH128	ABC	yes	yes	ZP_10870859.1
Sphingomonas sp. LH128	XKB1CXB2X	no	no	ZP_10874239.1
Sphingomonas sp. SKA58	ABC	yes	yes	ZP_01301928.1
Sphingopyxis alaskensis RB2256	ABC	no	yes	YP_617574.1
Sphingopyxis alaskensis RB2256	ABC	no	yes	YP_617642.1
Stenotrophomonas maltophilia R551-3	ABC	no	yes	YP_002028439.1
Sulfurovum sp. AR	DBKCA*	no	no	ZP_10062210.1
Sulfurovum sp. NBC37-1	ACKDXXB1B2	no	no	YP_001359375.1
Uncultured marine bacterium 463	KXXXCADBX	no	no	AAS07978.1
Xanthomonas axonopodis pv. *malvacearum* str. GSPB1386	ACB1B2DD	no	no	ZP_13082330.1
Xanthomonas axonopodis pv. *citri* str. 306	ACB1B2DD	no	no	XAC0058
Xanthomonas axonopodis pv. *malvacearum* str. GSPB1386	ACB1B2DD	no	no	ZP_13082330.1
Xanthomonas axonopodis pv. *malvacearum* str. GSPB2388	ACB1B2DD	no	no	ZP_13087953.1
Xanthomonas axonopodis pv. *punicae* str. LMG 859	ACB1B2DD	no	no	ZP_10261507.1
Xanthomonas citri pv. *magniferaeindicae*	ABC	yes	yes	ZP_09881129.1
Xanthomonas citri pv. *magniferaeindicae*	ACB1B2DD	no	no	ZP_09878744.1
Xanthomonas gardneri ATCC 19865	AABC	yes	yes	ZP_08185024.1

Supplementary Table S5a. Genome mining results II. Overview over putative lasso peptide precursors. Numbers on the right side represent the number of amino acids of the precursor peptides, the putative lasso peptides and the macrolactam rings, in this order. The threonine residue at the penultimate position of the leader peptide is highlighted in blue, while the amino acid at position 1 of the lasso peptide sequence and the corresponding ring forming amino acid are marked in red.

strain	precursor			
Asticcacaulis excentricus CB48 (2228)	MHTPIISETVQPKTAGLIVLGKASAE**T**R **G**LSQGVEP**D**IGQTYFEESRINQD	51	23	9
Asticcacaulis excentricus CB48 (2448)	MRTYNRSLPARAGLTDLGKVTTH**T**K **G**PTPMVGL**D**SVSGQYWDQHAPLAD	49	24	9
Asticcacaulis excentricus CB48 (2447)	MTKRTTIAARRVGLIDLGKATRQ**T**K **G**LTQIQAL**D**SVSGQFRDQLGLSAD	49	24	9
Breviundimonas diminuta ATCC 11568	MMNTFKLRLVSFGSAKAL**T**RD **G**MGEEFI-**E**GLVRDSLYPPAG	41	20	8
Burkholderia gladioli BSR3	MSAIIKKLFRNRASNQAHRFDVVSV**T**H **G**QPGYQTI**D**FRVVTRLGGR	46	19	9
Burkholderia mallei ATCC 23344	MVRFLAKLLRSTIHGSNGVSLDAVSS**T**H **G**TPGFQTP**D**ARVISRFGFN	47	19	9
Burkholderia oklahomensis C6786	MVRFLAKLLRSTIHGSNGVSLDAVSS**T**H **G**TPGFQTP**D**ARVISRFGFN	47	19	9
Burkholderia oklahomensis EO147	MVRFLAKLLRSTIHGSNGVSLDAVSS**T**H **G**TPGFQTP**D**ARVISRFGFN	47	19	9
Burkholderia pseudomallei 1106A	MVRFLAKLLRSTIHGSNGVSLDAVSS**T**H **G**TPGFQTP**D**ARVISRFGFN	47	19	9
Burkholderia pseudomallei 1710b	MVRFLAKLLRSTIHGSNGVSLDAVSS**T**H **G**TPGFQTP**D**ARVISRFGFN	47	19	9
Burkholderia pseudomallei 668	MVRFLAKLLRSTIHGSNGVSLDAVSS**T**H **G**TPGFQTP**D**ARVISRFGFN	47	19	9
Burkholderia pseudomallei 7894	MVRFLAKLLRSTIHGSNGVSLDAVSS**T**H **G**TPGFQTP**D**ARVISRFGFN	47	19	9
Burkholderia pseudomallei DM98	MVRFLAKLLRSTIHGSNGVSLDAVSS**T**H **G**TPGFQTP**D**ARVISRFGFN	47	19	9
Burkholderia pseudomallei K96243	MVRFLAKLLRSTIHGSNGVSLDAVSS**T**H **G**TPGFQTP**D**ARVISRFGFN	47	19	9
Burkholderia rhizoxinica HKI454	MNKQQQESGLLLAEESLMELCASSE**T**L **G**GAGQYK-**E**VEAGRWSDRIDSDDE	50	23	8
Burkholderia thailandensis E264	MVRLLAKLLRSTIHGSNGVSLDAVSS**T**H **G**TPGFQTP**D**ARVISRFGFN	47	19	9
Candidatus odyssella thessalonicensis L13(8821)	MKYLKKSIMVKSFFNKLDKNEHAGLINISQSASTI**T**A **G**GCGAKF-**E**QGGVTSLQKFDA	57	20	8
Candidatus odyssella thessalonicensis L13(8986)	MENIEITTKTSTVLDTNLVVVLEESASSL**T**Q **G**CCGGWL-**E**QWGGYSRP	47	16	8
Candidatus odyssella thessalonicensis L13(8991)	MENIEITTEISAILDTNLGVVLEESASSL**T**R **G**LVGTGK-**E**GFCVGFRY	47	16	8
Caulobacter crescentus CB15	MTPTTPRPTLLRLGAAKTL**T**R SLEDGTIK**E**AGSSQYYFV	39	18	9
Caulobacter segnis ATCC 21756 (2568)	MTKKNATQAPRLVRVGDAHRL**T**Q **G**AFVGQP-**E**AVNPLGREIQG	42	19	8
Caulobacter segnis ATCC 21756 (2567)	MTKTHRLIRLGDAQRL**T**Q **G**TLTPGLP**E**DFLPGHYMPG	37	19	8
Caulobacter segnis ATCC 21756 (2566)	MTSRFQLLRLGKADRL**T**R **G**ALVGLLL**E**DITVARYDPM	37	19	8
Caulobacter sp. AP07 (619)	MERTEHRIDDLIELGAASVE**T**K **G**ADLPTS-**E**VGIGRQPAGITED	43	21	8
Caulobacter sp. AP07 (620)	MSKETAMRQVENQTEELVELGRASDE**T**Q **G**PDMPTS-**E**VGIGRAPFGITED	49	21	8
Caulobacter sp. K31 (2238)	MNTLKTRLIRFGSAKRL**T**R AGTGVLLP**E**TNQIKRYDPA	38	19	9
Caulobacter sp. K31 (2239)	MTTPKFRLIRLGSAKRL**T**R SGIGDVFP**E**PNMVRRWD	36	17	9
Caulobacter sp. K31 (1984)	MTPIQSKFCLLRVGSAKRL**T**Q SFDVGTIK**E**GLVSQYYFA	39	18	9
Caulobacter sp. K31 (198x)	MTQVSPSPLRLIRVGRALDL**T**R SIGDSGLR**E**SMSSQTYWP	40	18	9
Caulobacter sp. K31 (5194)	MERIEDHIDDELIDLGAASVE**T**Q **G**DVLNAP-**E**PGIGREPTGLSRD	44	21	8
Caulobacter sp. K31 (5193)	MQRIIDETTDGLIELGAASVE**T**Q **G**DVLFAP-**E**PGVGRPPMGLSED	44	21	8
Caulobacter sp. K31 (519x)	MEFEGIPSPDARIDLGLASEE**T**C **G**QIYDHP-**E**VGIGAYGCEGLQR	44	21	8
Chitinophaga pinensis DSM 2588 (4903)	MKNETTQQKTTWETPDMTVLPINDV**T**L **G**AGAAGT-**D**FASEISV	42	15	8
Chitinophaga pinensis DSM 2588 (4902)	MKNNTSQQPLTWETPDMIVLPIDKV**T**L **G**KAGSGS-**D**FAAEWEVST	44	17	8
Citromicrobium bathyomarineum	MKEFDHNQNEVIDLGKASVE**T**K **G**AVGFYI-**D**ASGGQLANSPGLLDE	45	23	8
Citromicrobium sp. JLT1363	MNQPHETLPTNAGKRAWSTPQVVDLDDRRAIA **G**GAVNAP-**E**YFAPTFYSVS	50	18	8
Desulfobacca acetoxidans DSM 11109	MMICFPAVTRYLAIVLSLRRKASEL**T**WK QTPCRWT-**D**SRPSGTVP	44	16	8
Escherichia coli AY25 (Microcin J25)	MIKHFHFNKLSSGKKNNVPSPAKGVIQIKKSASQL**T**K **G**GAGHVP-**E**YFVGIGTPISFYG	58	21	8
Erythrobacter sp. NAP1	MTKVTEKVKYEAPKLTVFGSVRNL**T**G **G**SGTSFR-**D**GAVTNTRF	43	16	8
Geobacter uraniireducens Rf4	MDRDKQSCPDSSVISGESVAASTRSAWSAPTITRIEIKR**T**MA **G**ASVAL--**D**SSLFTTH	56	14	7

Supplementary Table S5b. Genome mining results II. Overview over putative lasso peptide precursors. Numbers on the right side represent the number of amino acids of the precursor peptides, the putative lasso peptides and the macrolactam rings, in this order. The threonine residue at the penultimate position of the leader peptide is highlighted in blue, while the amino acid at position 1 of the lasso peptide sequence and the corresponding ring forming amino acid are marked in red.

strain	precursor			
Haliangium ochraceum DSM 14365	MAKKKTDTKKTFLQDLGKATEE**T**K	SDGPPERY**D**NLVFRTKGPYLG	45	21 9
Legionella pneumophila 2300/99 Alcoy	MDKLDWELPEVCAIPIDSVTADQD	**G**NLKSDGP**D**SSGLSSSSSSSGG	46	22 9
Legionella pneumophila str. Corby	MDKLDWELPEVCAIPIDSVTADQY	**G**NKKNDGP**D**SSGLSSSSSPSG	45	21 9
Magnetospirillum magneticum AMB-1	MSDQSPPATEDAPARKPWHKPEITTLAVEET	ATNGSTGN**D**GSGATTFS	48	17 9
Marine gamma proteobacterium	MSGEQSVSRREENPQISISAGSGRKYKTPRLERFGEISHV**T**Q	**G**SFGKKK-**D**AEGDEVAIIPK	61	19 8
Marinomonas mediterranea MBP1	MSDIDVNTLLSEDKNSKKSLGDKLIWVAPVLEELETDK**T**RS	**G**GGTAT--**E**TPFNPGVSS	57	16 7
Mesorhizobium amorphae CCNWGS0123	MEQNVEKHEYETPSLTVHGSIETI**T**Q	**G**GGGSTAL**D**ASFPAHTPIGELTFS	50	24 9
Mesorhizobium australicum WSM2073	MEQNVEKHEYETPNLTVHGSIETI**T**Q	**G**GGGSTAI**D**ASFPAHTPIGDLTFS	50	24 9
Mesorhizobium ciceri biovar biserrulae	MEQNVEKHEYETPSLTVHGSIETI**T**Q	**G**GGGSTAI**D**ASFPAHTPIGDLTFS	50	24 9
Mesorhizobium loti MAFF303099	MEQNVEKHEYETPSLTVHGSIETI**T**Q	**G**GGGSTAI**D**ASFPAHTPIGDLTFS	50	24 9
Mesorhizobium opportunistum WSM2075	MEQNVEKHEYETPSLTVHGSIETI**T**Q	**G**GGGNTAI**D**ASFPAHTPIGDLTFS	50	24 9
Novosphingobium sp. AP12	MKEREHNESDVIELGIAHIE**T**R	**G**NALFGR**D**DTQTGMKYGVAGISSED	47	25 8/9
Novosphingobium sp. PP1Y (Lpi3xx)	MESRSRMEKIMEKTIDTLEDKVVDLGSVTEE**T**R	**G**IVGAL-**E**DQQGGQRAALGLTDD	55	22 7/8
Novosphingobium sp. PP1Y (AT163xx)	MKEVIMTVIDEHRDDELIDLGAVSAA**T**K	**G**AVQGEA-**D**LEGQPRQIFGAISDD	51	21 8
Novosphingobium sp. Rr 2-17	MDRLHDIIDLGAASVE**T**K	**G**PIGEGS-**D**LVLGQNPAGLSDD	39	19 8
Oligotropha carboxidovorans OM5	MQRVAGREGIHAVTRTRDAMKATARDP**T**I	**G**PLLIEH-**E**FENVRDGGRRGGDNQEMFGGRL	59	30 8
Phenylobacterium zucineum HLK1 (217x)	MNTRLHPPVSRLICLGGAKAS**T**N	QGVGIRQ-**E**LNPDEGFDD	40	17 8
Phenylobacterium zucineum HLK1 (239x)	MTRLLNLMSVRLLGFGSAKAA**T**N	**G**GIGGDF**E**DLNKPFDV	39	16 8/9
Pseudomonas geniculata N1	MFMNELIELGVVAGEQNHEDQSELIELGSVVSE**T**K	AHNHGTLW**D**GVFPTTSP	52	17 9
Rhodanobacter sp. 115	MTQTQEIVMNTNENIRNNASEDVIVLGVASVE**T**Q	**G**GPGSN--**E**AVGFGGPVIPGISEE	56	22 7
Rhodanobacter sp. 116-2	MTRPQETEMNTNENIRNDTPEDVIVLGVASVE**T**K	**G**PMGST**E**P**E**GGLGRPVIPGISEE	57	23 7/9
Rhodanobacter sp. 2APBS1	MNTNENIRTNAPEDVIELGIASVE**T**K	**G**EFGVT--**E**GGGQGKPMIPGISEE	48	22 7
Rhodanobacter thiooxydans LCS2	MTQSQETEMDTNENIRSNAQDDVIELSVASVE**T**K	**G**VLPIGN-**E**FMGHAATPGITE	54	20 8
Rhodobacter sphaeroides 2.4.1	MANNIQTLAYEAPVLRVHGTLEAM**T**H	**G**ATDGWAL**D**ASFPVDTPKGDLTFS	50	24 9
Rhodobacter sphaeroides ATCC 17025	MLRIFDVANNIQTLAYQAPVLRAHGTLEAM**T**H	**G**ASTGHVT**D**AAFPSGTAIADMTFS	56	24 9
Rhodobacter sphaeroides ATCC 17029	MANNIQTLAYEAPVLRVHGTLEAM**T**H	**G**ATDGWAL**D**ASFPVDTPKGDLTFS	50	24 9
Rhodobacter sphaeroides KD131	MLRVHGTLEAM**T**H	**G**ATDGWAL**D**ASFPVDTPKGDLTFS	37	24 9
Roseobacter denitrificans Och 114	MTKAVYEAPVLRSHGKVEAV**T**K	**G**GSTGSSL**D**AAFPAGTPFTDLTLS	46	24 9
Roseobacter litoralis Och 149	MTKAVYEAPVLRSHGKVEAV**T**K	**G**GSTGSSL**D**AAFPAGTPFADLTLS	46	24 9
Rubrivivax benzoalyticus	MDEETELDIVDLGDAKDT**T**K	**G**ILLPGS**E**DNVTMPGRQIA	39	19 8/9
Rubrivivax gelatinosus IL144	MKEFAMDEELELEIVDLGDAKEL**T**Q	**G**APSLINS**E**DNPAFPQRV	43	18 9
Shewanella sp. HN-41	MQQDNMNKSGLAVEQTVKAAWSKPEIEEINIAGY**T**E	**G**KATGTN-**E**SFLSAPS	51	15 8
Sphingobium chlorophenolicum L-1	MHIPYVKSGREPGRPLVKLGKARHQ**T**R	ASMNEIAP**E**LVGDKTQRFGG	47	20 9
Sphingobium japonicum UT26 (C1-3001x)	MERDNDVIELGAVSVE**T**K	**G**PGGITG-**D**VGLGENNFGLSDD	39	21 8
Sphingobium japonicum UT26 (P1-00960)	MDRHDNSEVDEIIDLGTASAV**T**Q	**G**MGSGST-**D**QNGQPKNLIGGISDD	46	23 8
Sphingobium sp. AP49	MERDAHSQDIVDLGDAVEL**T**R	**G**VLLIGQ**D**DGNGTLYKASGISDED	44	24 8/9
Sphingobium sp. SYK-6	MNHEHETDDLIELGNASTD**T**K	**G**GAFIVE-**D**NEGSLSLQTGLTDD	43	22 8
Sphingobium yanoikuyae ATCC 51230 (4094)	MEREEEIMRIDLGAATEQ**T**Q	**G**PGGIDG**D**EFLQQDRVGLADD	41	21 8/9
Sphingobium yanoikuyae ATCC 51230 (4584)	MEKESYESPELVELGSFESL**T**Q	**G**TSYGESL**D**ATFPDGTPRGELTFS	46	24 9
Sphingobium yanoikuyae XLDN2-5 (284x)	MERNSEDRRDDVVELGAVSVE**T**K	**G**ISGGTV-**D**APAGQGLAGILDD	44	21 8
Sphingobium yanoikuyae XLDN2-5 (616x)	MEREYEATDDGVVELGVGSEL**T**N	**G**VGRGGI-**D**SPDRQPLTGLLDD	44	21 8

Supplementary Table S5c. Genome mining results II. Overview over putative lasso peptide precursors. Numbers on the right side represent the number of amino acids of the precursor peptides, the putative lasso peptides and the macrolactam rings, in this order. The threonine residue at the penultimate position of the leader peptide is highlighted in blue, while the amino acid at position 1 of the lasso peptide sequence and the corresponding ring forming amino acid are marked in red.

strain	precursor			
Sphingobium yanoikuyae XLDN2-5 (3348)	MEKESYESPELVELGSFESL**T**Q **G**TSYGESL**D**ATFPDGTPRGELTFS	46	24	9
Sphingomonas sp. LH128	MSQQRRNVMEHVEEDTMVELGAVSTE**T**K **G**GQVGIN-**D**FERGLQLTGSGLTDD	51	23	8
Sphingomonas sp. SKA58	MEMKMNNINEHEDSVVDLGVASVE**T**K **G**AALDDS-**D**NIGGQVRQLGIADD	48	22	8
Sphingopyxis alaskensis RB2256 (253x)	MKDFNELIDLGAISVE**T**R **G**IEPLGPV**D**EDQGEHYLFAGGITADD	44	26	9
Sphingopyxis alaskensis RB2256 (260x)	MERTEVIEEVIDLGKASVE**T**K **G**EALIDQ-**D**VGGGRQQFLTGIAQD	44	23	8
Stenotrophomonas maltophilia R551-3	MNELIELGVVAGQQNPEDQGVGGELIELGSVVSE**T**K AHNHGTLW**D**GVFPTTSP	53	17	9
Sulfurovum sp. AR	MRVRMKKNFSIKKYNKPEIKNLGKIGTKTQYNKSTAGADG**T**S STHYNKGQ**D**TGFGGAGTVTP	62	20	9
Sulforovum sp. NBC37-1	MIEKNHSDFSGETERQKKPYSSPKIEIIANVRMA**T**K **G**STIGGT-**D**SGGQSANVPSDPGGW	59	23	8
Uncultured marine bacterium	MSGEQSVSRREENPQISISAGSGRKYKTPRLERFGEISHV**T**Q **G**SFGKKK-**D**AEGDEVAIIPK	61	19	8
Xanthomonas axonopodis pv. *citri* str..306	MHENQPAVLSDTSIEGGRKAWSAPVVSFLSIDE**T**A SNATVG-**DD**GNGTFTGS	51	16	7/8
Xanthomonas axonopodis pv. malvercearum str. GSPB1386	MHENQPAVLSDTSIEGGRKAWSAPVVSFLSIDE**T**A NNATVG-**DD**GNGTFTGS	51	16	7/8
Xanthomonas axonopodis pv. malvercearum str. GSPB2388	MHENQPAVLSDTSIEGGRKAWSAPVVSFLSIDE**T**A NNATVG-**DD**GNGTFTGS	51	16	7/8
Xanthomonas axonopodis pv. punicae str. LMG 859	MHENQPAVLSDTSIEGGRKAWSAPVVSFLSIDE**T**A SNATVG-**DD**GNGTFTGS	51	16	7/8
Xanthomonas citri pv. *mangiferaindicae* (260x)	MKTFVDSAPSQEDHGKDIIVLGIASVE**T**Q **G**GAGAG--**E**VNGMSPIAGISEE	49	20	7
Xanthomonas citri pv. *mangiferaindicae* (18x)	MHENQPAVLSDTSIEGGRKAWSAPVVSFLSIDE**T**A SNATVG-**DD**GNGTFTGS	51	16	7/8
Xanthomonas gardneri ATCC 19865 (4058)	MNSNDTTHSDASNEITVLGVASTD**T**K **G**GPLAG-**EE**IGGFNVPGISEE	46	20	7/8
Xanthomonas gardneri ATCC 19865 (4059)	MDTSNNDARTTALDQDLIVLGVASLD**T**Q **G**GPLAG-**EE**MGGITTLGISQD	48	20	7/8

Supplementary Table S6. Comparison of the processing enzymes from different gene clusters of the same organism with ABC organization and promising precursor peptides. Identity and similarity were determined using the BLOSUM62 algorithm.

	identity / similarity B	identity / similarity C
S. japonicum UT26	37% / 52%	34% / 49%
S. yanoikuyae XLDN2-5	33% / 49%	-
S. alaskensis RB2256	34% / 49%	33% / 48%

Supplementary Table S7. Production values of the investigated lasso peptides using the WT constructs or the modified constructs carrying an *E. coli* optimized RBS.

construct / fermentation condition	UV peak area (main product)	production increase through RBS incorporation
burhABC pET41a / 1 d 37 °C	41399295[a] (-6 aa)	-
burhA_RBS_BC pET41a / 1 d 37 °C	513371242[a] (-6 aa)	12.4-fold
ck31_A1A2A3BC pET41a / 3 d 20 °C	36664950 (only caulonodin I, -4 aa)	-
ck31_A1_RBS_BC pET41a / 3 d 20 °C	3099721597 (-4 aa)	84.5-fold
ck31_A2_RBS_BC pET41a / 3 d 20 °C	2594419029 (-4 aa)	nd
ck31_A3_RBS_BC pET41a / 3 d 20 °C	158874664 (-4 aa)	nd
pzucABC pET41a / 3 d 20 °C	-	-
pzucA_RBS_BC pET41a / 3 d 20 °C	156292239[a] (full length)	nd
rhotABC pET41a / 3 d 20 °C	6976504 (-3 aa)	-
rhotA_RBS_BC pET41a / 3 d 20 °C	5087403 (-3 aa)	no increase
rugeABC pET41a / 1 d 37 °C	53597497 (full length)	-
rugeA_RBS_BC pET41a / 1 d 37 °C	1339158610 (full length)	25.0-fold
sjap1_ABC pET41a / 3 d 20 °C	166128796 (-4 aa)	-
sjap1_RBS_BC pET41a / 3 d 20 °C	295433797 (-4 aa)	1.8-fold
sjap2_ABC pET41a / 3 d 20 °C	35147323 (-4 aa)	-
sjap2_RBS_BC pET41a / 3 d 20 °C	85936911 (-4 aa)	2.5-fold
syan1_ABC pET41a / 3 d 20 °C	187078501 (-4 aa + -5 aa)	-
syan1_RBS_BC pET41a / 3 d 20 °C	1664150210 (-4 aa + -5 aa)	8.9-fold
sala1_ABC pET41a / 3 d 20 °C	242675653 (-5 aa)	-
sala1_RBS_BC pET41a / 3 d 20 °C	1851952312 (-5 aa)	7.6-fold
sala2_ABC pET41a / 3 d 20 °C	18517684 (-4 aa)	-
sala2_RBS_BC pET41a / 3 d 20 °C	265940664 (-4 aa)	14.4-fold

nd = not determined

[a]as no UV peak was observable, the peak area of the EIC of the respective singly and doubly charged ions of the investigated compound was used instead.

Supplementary Table S8. Alignment of the first motif of the B proteins from the 15 known, functional lasso peptide biosynthetic gene clusters identified by use of the MEME algorithm.

protein	position of the first aligned amino acid	sequence	position of the last aligned amino acid
AtxB	2	YELNDGVGLALVDQHPIFLDLKTDRYLSLSP	32
BurhB	33	HHLMPHVFGACFAQQIILLDVVRDRYAVLTE	63
CapB	1	MTPASHCHIAVFDQAIVALDMQRSRYFLYDE	31
CK31_B	3	VAVPDHLAYCVKQGGVTFLDVRGDRYFGLPP	33
CsegB	3	LWLSAGVYAVMIDDDVVFLDVATNAYFCLPA	33
LarB1	3	LRLRKNVIITPTEYGAVALDERSGDYYQLNS	33
PzucB	2	PALHPKSFLVAVDADLVLLQVDRDRYLCLPD	32
RhotB	5	YWLHDDLSFCLVDGHPIFLDMGSDRYFRLSG	35
RugeB	12	LSLAPHVRACATGGQVILLDLRSGTYLGLAG	42
Sjap1_B	4	FTLRDGLSFCRTEGQVIFLDLHADRYFGLRP	34
Sjap2_B	3	LALRDQLSFCIIEQHVLFLDVANDRYFGVRE	33
SyanB	1	MA.RADINWCDVGGRIVVIDVARDIYLQLPQ	30
Sala1_B	3	AMLRSGLHYCQFDSGFVLFDLSADRYFLLTD	33
Sala2_B	6	YRTAPGIFYCRAGSRVIFLDMNRDRYFQLGA	36

Supplementary Table S9. Alignment of the second motif of the B proteins from the 15 known, functional lasso peptide biosynthetic gene clusters identified by use of the MEME algorithm.

protein	position of the first aligned amino acid	sequence	position of the last aligned amino acid
AtxB	152	CLPDAFAIATHLRRR.GVDAKLVFGV..R.LPFAAHAWVQVDDIVVGDRPDRI	200
BurhB	209	CLEWSVAL.TVLCARAGLALKLVIGV..QSFPFYAHAWSEAGGAVIGDSPLRR	258
CapB	169	CLQMSFALATHLRRE.NVPAQLVIGV..RPMPFVAHAWVEIDGRVCGDEPELK	218
CK31_B	161	CLLDSLALLDFLHRR.GLYPHIVFGVI.R.QPFAAHCWVQADDVVLNDRLDHV	210
CsegB	158	CLFRSQMLRDYLLAL.GHRVDWIFGV..RTWPFGAHCWLQAGDLVLDDEAERL	207
LarB2	80	CLQRSISV.CILMRLDGRWPTWCVGVPSK.PPFRAHAWIEAGGQIVA.ELGDM	129
McjB	150	CLTYSYALKRILNSR.NIDAHLVIGV..RTQPFYSHSWVEVGGQVINDAPNMR	199
PzucB	154	CLVRSFVLRRFLRR.SGVDADWVFGV..RTWPFSAHCWLQVGDRALDDYAERL	203
RhotB	164	CLLDSLAMAKFLKRR.RMRVNVIVGV..TGDPFSAHCWVQAGDMVLNDALGNV	213
RugeB	183	CLYDSLALLLFLASE.GHCPRWVVGV..RTRPFAAHSWVQSGSLVLNDQAEHV	232
Sjap1_B	160	CLPTALSLASWLIAL.DIRPDLVLGV..KLNPFQAHSWVEVDGMIVADDPDQV	209
Sjap2_B	159	CLLYSLALAKLLLSS.GTSPTLIFGV..KLAPFAAHCWVQDGDCVLNDDLDHT	208
SyanB	162	CLPDSMALLHFLRSR.NMACRLVLGV..RLDPFEAHCWVQDHDMILGDTHERV	211
Sala1_B	156	CLVRGIAMARMLGCR.GVAASLVFGV..T.MPFAAHSWVQVGDTVLTDSLDMV	204
Sala2_B	161	CLSRSIALVQRLAAQ.GCAAHLVIGV..RVIPFAAHSWAQIGDVVLNDSPEEV	210

Supplementary Table S10. Alignment of the first motif of the C proteins from the 15 known, functional lasso peptide biosynthetic gene clusters identified by use of the MEME algorithm.

protein	position of the first aligned amino acid	sequence	position of the last aligned amino acid
AtxC	204	ELRALILRSIKAYASDFPHVVVSFSGGLDSSVVAAGLAQ	242
BurhC	210	TPFDILKHYLNRFVSFSDRPVLELSGGLESSSVLLAMRA	248
CapC	204	SCAAALDDVYSRIAHSHPNVCAALSGGVDSSAGAIFLRK	242
CK31_C	215	AVRQAVEVSVRKWADQSSPVLLELSGGLDSSIIACCLDE	253
CsegC	221	DLRRTIDATVAALASDAGPIVCEISGGLDSAIVATSLAA	259
LarC	209	ETAAALFSVVRNQLSDHSAASCDISGGLDSSSIAAIAAN	247
McjC	176	STIDSIIDNI.EMMRDNRKIALLFSGGLDSALIFHTLKE	213
PzucC	204	ALAASLDDVCATWTGAYGALMAEVSGGFDSALVAVTARR	242
RhotC	221	AIRSSVETVVKSLADLDQLTLLELSGGLDSSIIGACLKR	259
RugeC	227	ALRETVQACASAWAGCHDRVLFRLSGGVDSSILLSCLSR	265
Sjap1_C	216	MIRRAVSTATRASTMGTKRALLMLSGGIDSSVLAAALGT	254
Sjap2_C	216	RVRRVTLNSVHAWASCYRNILIQISGGLDSSILAAALRG	254
SyanC	218	EVAQAVDSACHHWSRQVGPVALTLSGGLDSSVLAAALPG	256
Sala1_C	203	ALRRTIADCVGAWATCFPSIMVGVSGGLDSSIVAASAAN	241
Sala2_C	203	RLRDTVFASVGALASDYQNIIVGVSGGLDSSIVCAALTR	241

Supplementary Table S11. Alignment of the second motif of the C proteins from the 15 known, functional lasso peptide biosynthetic gene clusters identified by use of the MEME algorithm.

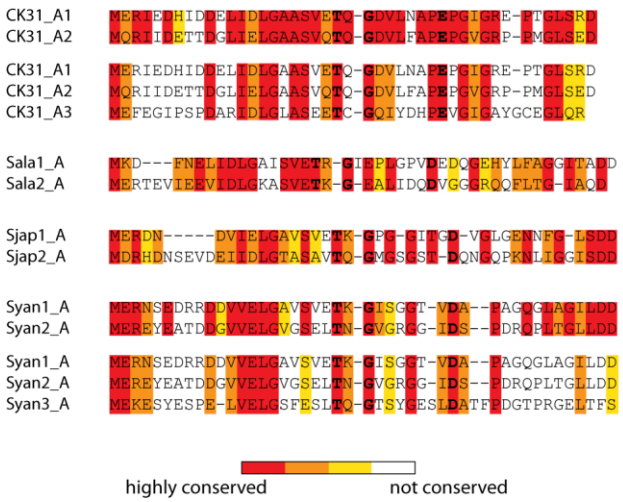
protein	position of the first aligned amino acid	sequence	position of the last aligned amino acid
AtxC	316	S.QTRT.G.GAI.FSGNGGDSVFCFMH..SA	340
BurhC	324	SQFERQEH.TPL.ISGHGGDALFLAPPPCSA	352
CapC	319	A.LGPS...SVL.LEGQGGDLLFRAVP..DA	342
CK31_C	324	AELAPQLGARSF.FSGLGGDNVFCSIA..TA	351
CsegC	334	A.LEAI.DARAL.FTGHGGDTVFYQVAA.SA	360
LarC	329	N.VARSKGSQVH.LTGFGGDELFIGSP..T.	354
McjC	285	G.QPPIDD.DLLYLCGHGGDHIFGQNP..S.	310
PzucC	318	GELNRHK.ADAL.LTGYGGDAVFFQMP..SA	344
RhotC	332	S.AARCE.ANSF.FSGAGGDAVFGYLT..SA	357
RugeC	337	AELAAEYGSGAL.FTGAGGDQLFFEFG..R.	363
Sjap1_C	327	A.VGAS.GADLS.IDGGGGDNVFFSLR..S.	351
Sjap2_C	327	A.AKAV.GAEVV.FNGNAGDAVFCFLQ..SA	352
SyanC	325	A.VAHEVGARAI.AYGSGGDNVFCFLR..SA	351
Sala1_C	316	G.QLPV...DAF.FSGNGGDGILCGVR..SA	339
Sala2_C	313	AEICKT.GADAI.FTGNGGDNVFCFLH..SA	339

Supplementary Table S12. Alignment of the third motif of the C proteins from the 15 known, functional lasso peptide biosynthetic gene clusters identified by use of the MEME algorithm.

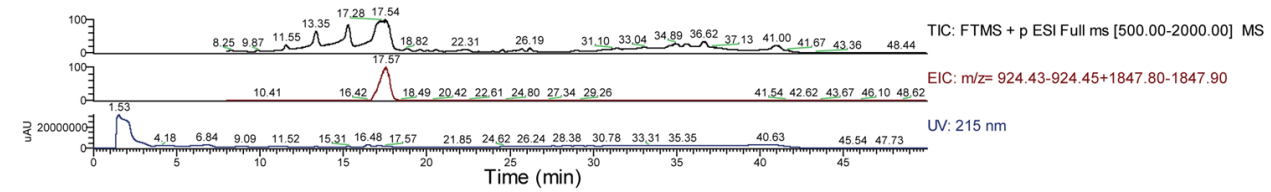
protein	position of the first aligned amino acid	sequence	position of the last aligned amino acid
AtxC	444	P.PVVHPLMAKPIQAFCLSLPSWMWVSGGKDRSLVRDAFEGLLP	486
BurhC	460	P.KH.YPFLCQPMVEFALSTPSYDHFEGAHNRIVLRKSVAAATG	501
CapC	451	S.RL.NPYLAQPVVEAAFGLHSYDSFDHRNDRIVLREIASAHTP	492
CK31_C	458	AVRF..PLLTQPVMEACLRVPTWMANHQGRNRAVARDAFFDRLP	499
CsegC	462	A.RIVHPLLAQPVVEACLAIPAPILSAGEGERSFAREAFADRLP	504
LarC	482	GVLLVAPMLEQAVVEAAISVRTPERLTPHKYKPVLTHATRGLLP	525
PzucC	448	A.DLLYPLLSQPLVETALAIPSWKHLDGGRDRALARSVAADRLP	490
RhotC	461	PIRM..PLLSQPVIETCLRVPSWMWFHNGQNRAVARQAFSDLLP	502
RugeC	473	P.ELVNPLLSQPLIELCLRLPTWVLTHGGRGRALARRAFADELP	515
Sjap1_C	459	PVRHVTPLASQPVVEACLRIPSWWWFRQGQNRVVARDAFRDRLP	502
Sjap2_C	458	PIFMVSPLLSQPLVELCLAVPSWMWCAGGVNRALARAAFERDLP	501
SyanC	458	GLPVIYPLLSQPVMEACLRQPTWLWYEDGMNRAIARRAYAPKLP	501
Sala1_C	448	P.PHIAPLMSQPIIEACLAIPTWDWIAGGQNRAVARKAFEGCLP	490
Sala2_C	447	P.PMFPVLLAQPIVETCLRISSWQWCEGGINRSVARRAFGGFLP	489

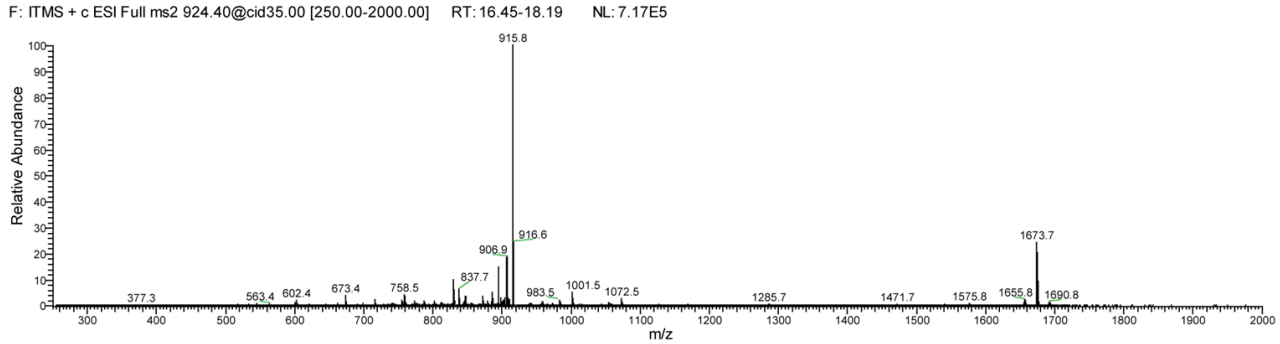
Supplementary Table S13. Alignment of the fourth motif of the C proteins from the 15 known, functional lasso peptide biosynthetic gene clusters identified by use of the MEME algorithm.

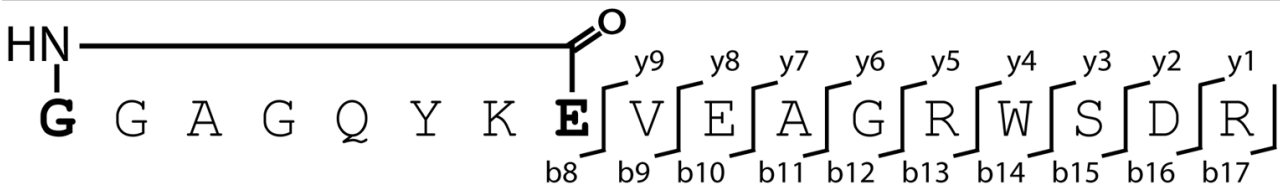
protein	position of the first aligned amino acid	sequence	position of the last aligned amino acid
AtxC	480	AFEGLLPDSVRLRKSKGSPAGFLHALYRAKGRQMIERIRHGYLRREGIID	529
BurhC	498	AATGY.PN..LWRRNKGETSGIDLLGLRDHREHVMAVCLEGWLAKAGYID	544
CapC	487	ASAHT.PVDVLWRRTKGSFGIGFVKGIVSHYDALRELIRDGVLMRSGRLD	535
CK31_C	493	AFFDRLPPRVRDRQTKGGLNAFMGVAFERNRQALARHLLDGRLVQRGLID	542
CsegC	498	AFADRLPPSIVGRRSKGEISVFLNRSLAASAPFLRGFLLEGRLAARGLID	547
LarC	519	ATRGLLPAVVAERQTKGGEDTDAAIGFSENISAIRELWDESRLASLGIVD	568
PzucC	485	A.ADRLPPEIVARRSKGMLTSYYARQVCESLGFLRAYLLDGRLAAEGVLD	533
RhotC	496	AFSDLLPQKIIARKSKGTLTAYLGALHRRRKHEITDFLADGQLQAHGLVD	545
RugeC	509	AFADELPPEIRTRRAKGGMEELLQAVLLRNLGFAKELLLDGELARRGLLD	558
Sjap1_C	496	AFRDRLPAAVVDRRTKGTPDSYIAEIFTANRSVIRSMLLDGNLRRAGLLD	545
Sjap2_C	495	AFERDLPRQIIQRRSKGSPDSLLLAILELNRPALIKWLCDGLLASRNIID	544
SyanC	495	AYAPKLPARIVQRASKGGFTSLVRSLYLRNLAAIRAMLLEGELCARGILD	544
Sala1_C	484	AFEGCLPELITARTQKGGPGDFHLSIYRAHRSALHDRLRGGILAESGILD	533
Sala2_C	483	AFGGFLPRQVVHRTSKAGPDSVTAAAFESGRPKLRELLLDGLLRQNRIID	532



Supplementary Figure S1 - Comparison of suitable precursor peptides found in the same organism and/or in the same gene cluster. In cases of the precursor peptides found in *C.* sp. K31 and *S. yanoikuyae* XLDN2-5 two alignments are shown, one for the highly similar precursor peptides and one for all that were compared from these organisms.

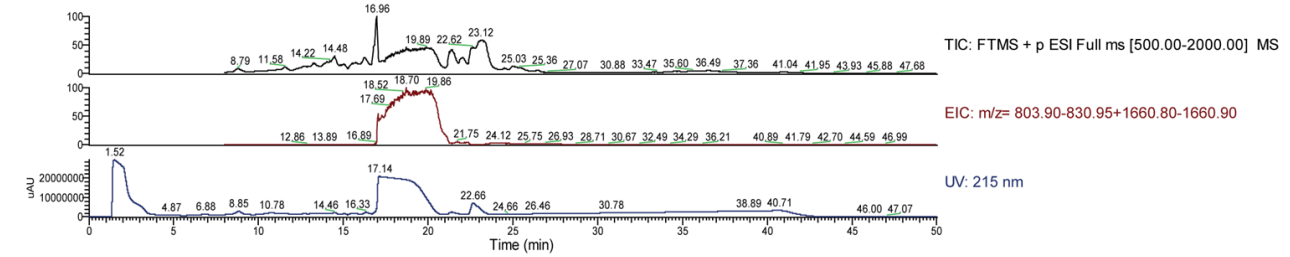


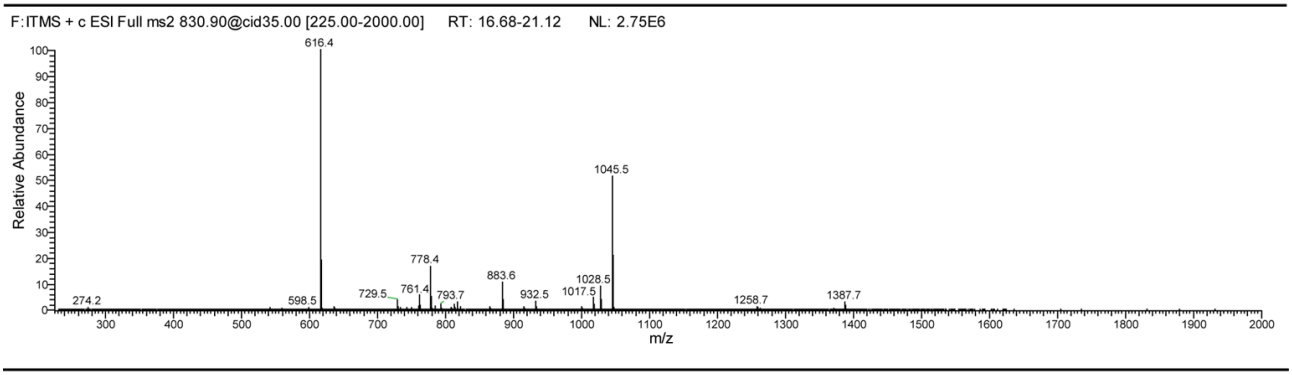


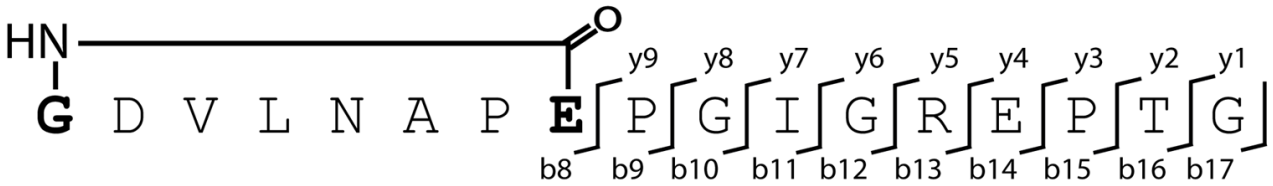


fragment ion	charge	predicted mass	observed mass	intensity	fragment ion	charge	predicted mass	observed mass	intensity
b17	+2	915.4	915.8	717000	**b10-H_2O**	+1	983.5	983.5	14800
b17-H_2O	+2	906.4	906.9	134000	**b10-2H_2O**	+1	965.5	965.5	3780
b16	+2	837.4	837.7	45900	**b9**	+1	872.4	872.5	24400
b16	+1	1673.8	1673.7	146000	**b8**	+1	773.4	773.5	13100
c15	+1	1575.8	1575.8	4230	**b8-H_2O**	+1	755.4	755.4	13900
b15	+1	1558.7	1558.8	1030	**y2**	+1	290.2	290.2	693
b15-H_2O	+1	1540.7	1540.8	1930	**y3-H_2O**	+1	359.2	359.3	1030
b14	+1	1471.7	1471.7	2350	**y3**	+1	377.2	377.3	1590
b14-H_2O	+1	1453.7	1453.7	654	**y4-H_2O**	+1	545.3	545.4	5070
b13	+1	1285.6	1285.7	4100	**y4**	+1	563.3	563.4	5660
b13-H_2O	+1	1267.6	1267.7	906	**y6-H_2O**	+1	758.4	758.5	28700
b12	+1	1129.5	1129.6	565	**y6**	+1	776.4	776.5	5500
b11	+1	1072.5	1072.5	18000	**y7**	+1	847.4	847.5	24300
b11-H_2O	+1	1054.5	1054.5	6680	**y8**	+1	976.5	976.6	2030
b10	+1	1001.5	1001.5	36500	**y9**	+1	1075.5	1075.6	995

Supplementary Figure S2 - Mass analysis of burhizin. On top the total ion current (TIC), extracted ion current (EIC) for the singly and doubly charged ions of the main product (-6 aa truncation) and the UV signal of the 1 d 37 °C pellet extract are shown. Depicted in the middle is the fragmentation spectrum of the doubly charged main product. All detectable fragment ions with their predicted masses, observed masses and signal intensities are listed on the bottom.

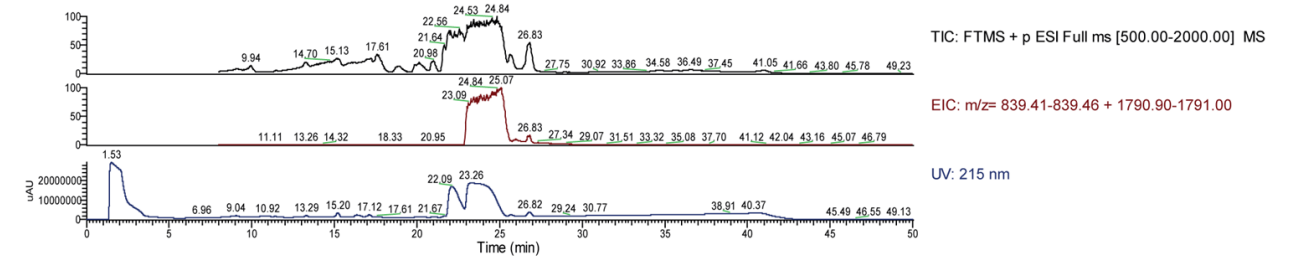


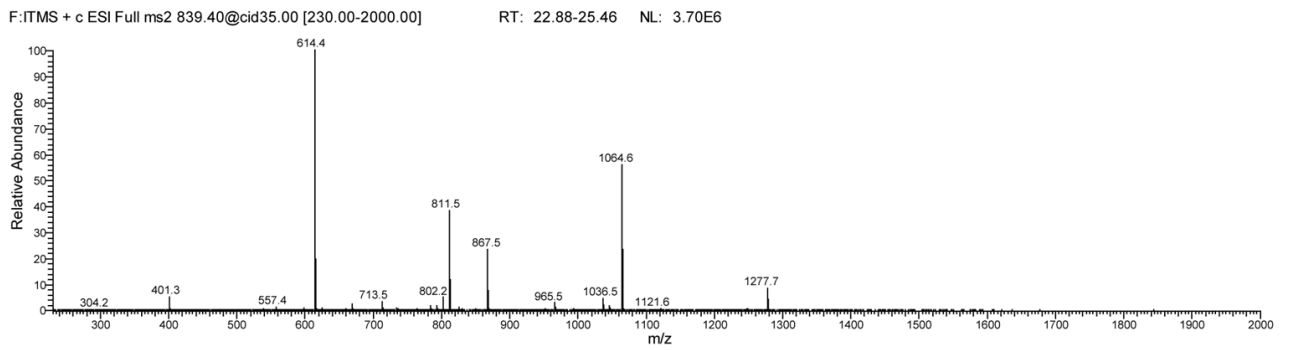


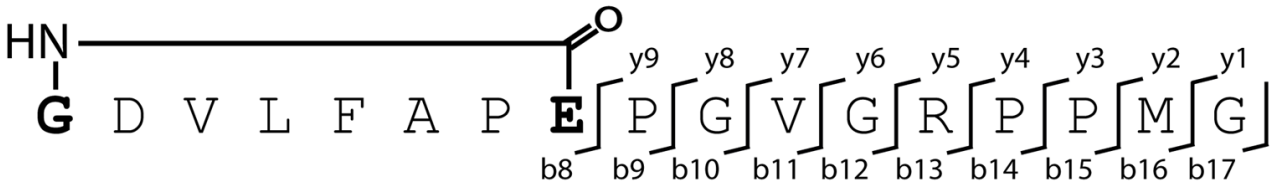


fragment ion	charge	predicted mass	observed mass	intensity	fragment ion	charge	predicted mass	observed mass	intensity
b17-NH_3	+2	813.4	813.5	52800	**b10-NH_3**	+1	915.4	915.5	26900
b16	+2	793.4	793.7	67200	**b8**	+1	778.4	778.4	459000
b14	+2	694.4	694.7	8420	**b8-NH_3**	+1	761.4	761.4	159000
b14	+1	1387.7	1387.7	76700	**y3-H_2O**	+1	256.1	256.2	1460
b14-NH_3	+1	1370.7	1370.7	11700	**y3**	+1	274.1	274.2	17000
b13	+1	1258.7	1258.7	30200	**y4-H_2O**	+1	385.2	385.3	2630
b13-NH_3	+1	1241.6	1241.8	3160	**y4**	+1	403.2	403.3	6760
b12	+1	1102.6	1102.6	5150	**y5-H_2O**	+1	541.3	541.4	21300
b12-NH_3	+1	1085.5	1085.6	2070	**y5**	+1	559.3	559.4	13300
b11	+1	1045.5	1045.5	1420000	**y6-H_2O**	+1	598.3	598.5	22000
b11-NH_3	+1	1028.5	1028.5	250000	**y6**	+1	616.3	616.4	2750000
a11	+1	1017.5	1017.5	128000	**y7**	+1	729.4	729.5	101000
a11-NH_3	+1	1000.5	1000.5	32300	**y9-H_2O**	+1	865.5	865.6	26200
b10	+1	932.5	932.5	86800	**y9**	+1	883.5	883.6	294000

Supplementary Figure S3 - Mass analysis of caulonodin I. On top the total ion current (TIC), extracted ion current (EIC) for the singly and doubly charged ions of the main product (-4 aa truncation) and the UV signal of the 3 d 20 °C pellet extract are shown. Depicted in the middle is the fragmentation spectrum of the doubly charged main product. All detectable fragment ions with their predicted masses, observed masses and signal intensities are listed on the bottom.

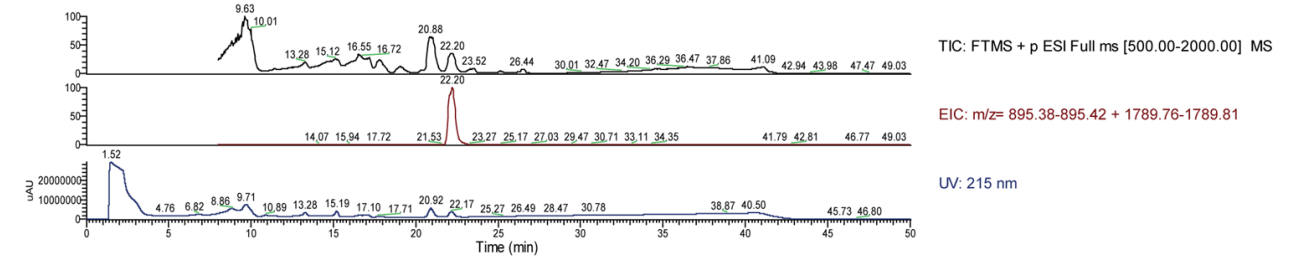


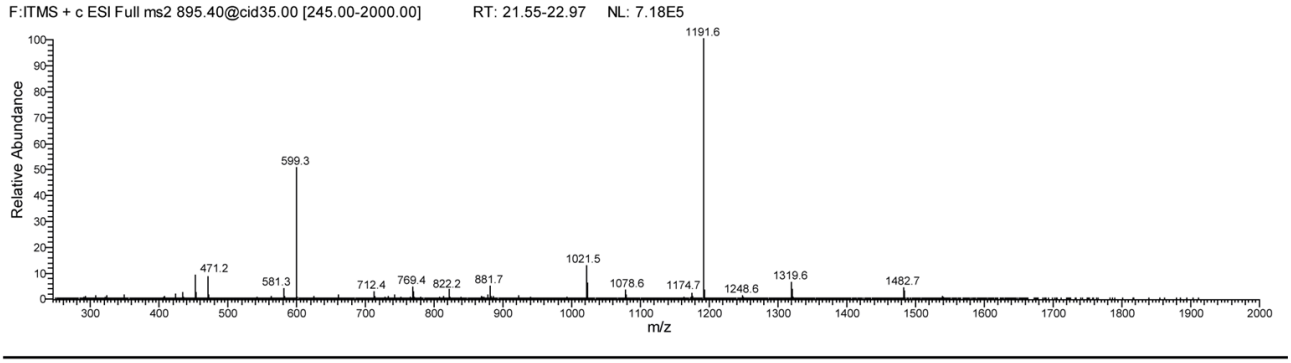


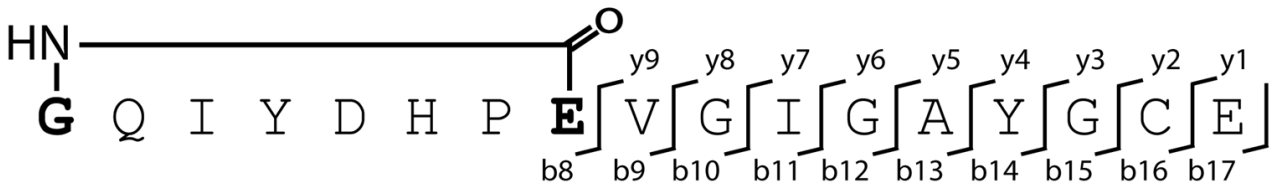


fragment ion	charge	predicted mass	observed mass	intensity	fragment ion	charge	predicted mass	observed mass	intensity
b17	+2	830.4	830.6	22100	**b10**	+1	965.5	965.5	108000
b17-H_2O	+2	821.4	821.8	11000	**b9**	+1	908.5	908.5	5510
b16	+2	801.9	802.2	178000	**b8**	+1	811.4	811.5	1390000
b15	+2	736.4	736.7	17100	**b8-H_2O**	+1	793.4	793.5	66000
b14	+2	687.9	688.2	10300	**y3**	+1	304.1	304.2	9940
b14	+1	1374.8	1374.8	3170	**y4**	+1	401.2	401.3	181000
b13	+1	1277.7	1277.7	300000	**y5-H_2O**	+1	539.3	539.4	18400
b13-H_2O	+1	1259.7	1259.7	7770	**y5**	+1	557.3	557.4	39700
b12	+1	1121.6	1121.6	21300	**y6**	+1	614.3	614.4	3700000
b12-H_2O	+1	1103.6	1103.6	2560	**y7**	+1	713.4	713.5	114000
b11	+1	1064.6	1064.6	2030000	**y8**	+1	770.4	770.6	9870
b11-H_2O	+1	1046.5	1046.6	57000	**y9**	+1	867.5	867.5	849000

Supplementary Figure S4 - Mass analysis of caulonodin II. On top the total ion current (TIC), extracted ion current (EIC) for the singly and doubly charged ions of the main product (-4 aa truncation) and the UV signal of the 3 d 20 °C pellet extract are shown. Depicted in the middle is the fragmentation spectrum of the doubly charged main product. All detectable fragment ions with their predicted masses, observed masses and signal intensities are listed on the bottom.

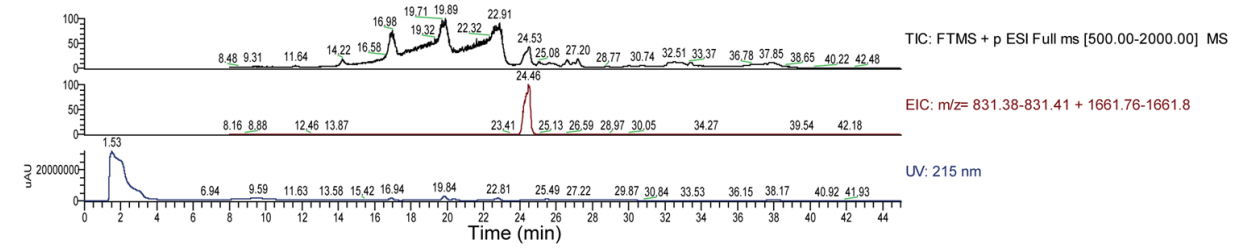


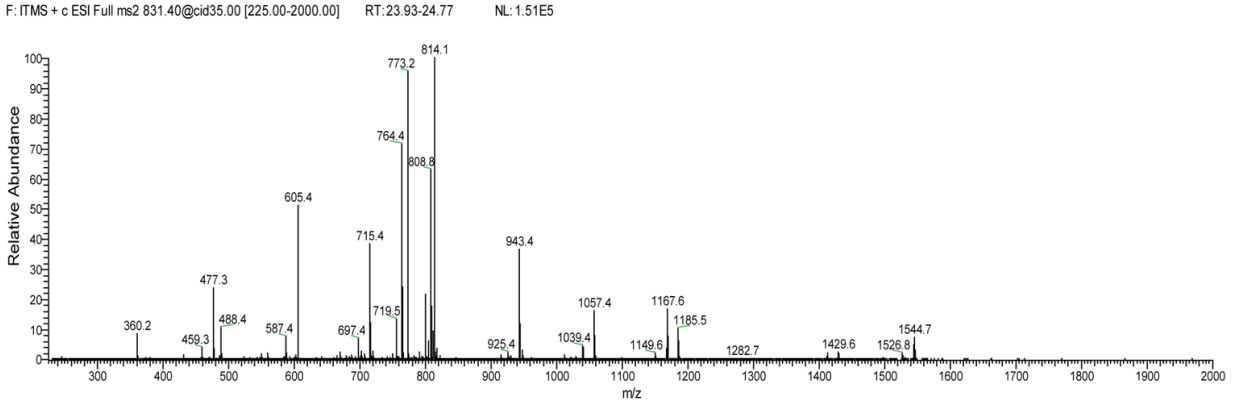


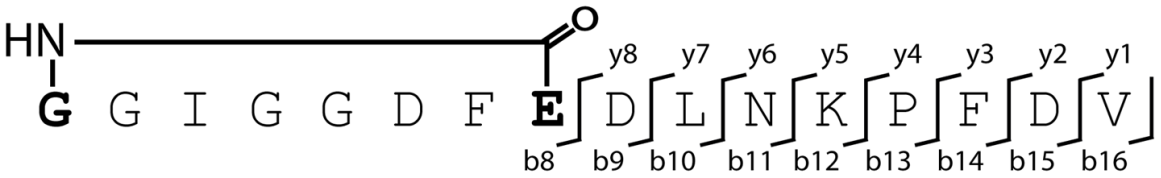


fragment ion	charge	predicted mass	observed mass	intensity	fragment ion	charge	predicted mass	observed mass	intensity
b17	+2	886.4	886.7	5550	**b10**	+1	1078.5	1078.6	24600
b16	+2	821.9	822.2	25600	**b9**	+1	1021.5	1021.5	94500
b16-H_2O	+2	812.9	813.2	5990	**a9**	+1	993.5	993.6	3490
b14	+2	741.9	742.1	9000	**b8**	+1	922.4	922.5	7810
b13	+2	660.3	660.5	9330	**y2**	+1	251.1	251.1	1040
b12	+2	624.8	625.0	9520	**y3**	+1	308.1	308.1	7660
b16	+1	1642.7	1642.7	1940	**y4**	+1	471.2	471.2	61300
b15	+1	1539.7	1539.7	5310	**y5**	+1	542.2	542.3	2900
b14	+1	1482.7	1482.7	30300	**y6-H_2O**	+1	581.2	581.3	27500
b13	+1	1319.6	1319.7	46900	**y6**	+1	599.2	599.3	372000
b12	+1	1248.6	1248.6	8310	**y7**	+1	712.3	712.4	18200
b11	+1	1191.6	1191.6	718000	**y8**	+1	769.3	769.4	32000
b11-NH_3	+1	1174.6	1174.7	14000	**y9**	+1	868.4	868.5	5580
a11	+1	1163.6	1163.7	2640					

Supplementary Figure S5 - Mass analysis of caulonodin III. On top the total ion current (TIC), extracted ion current (EIC) for the singly and doubly charged ions of the main product (-4 aa truncation) and the UV signal of the 3 d 20 °C pellet extract are shown. Depicted in the middle is the fragmentation spectrum of the doubly charged main product. All detectable fragment ions with their predicted masses, observed masses and signal intensities are listed on the bottom.

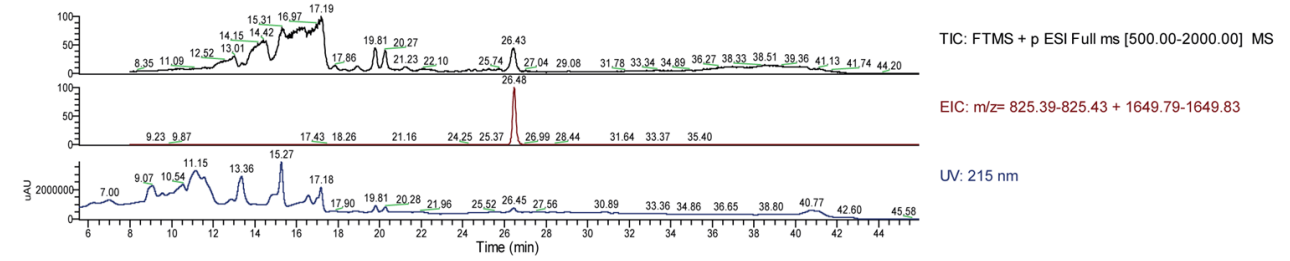


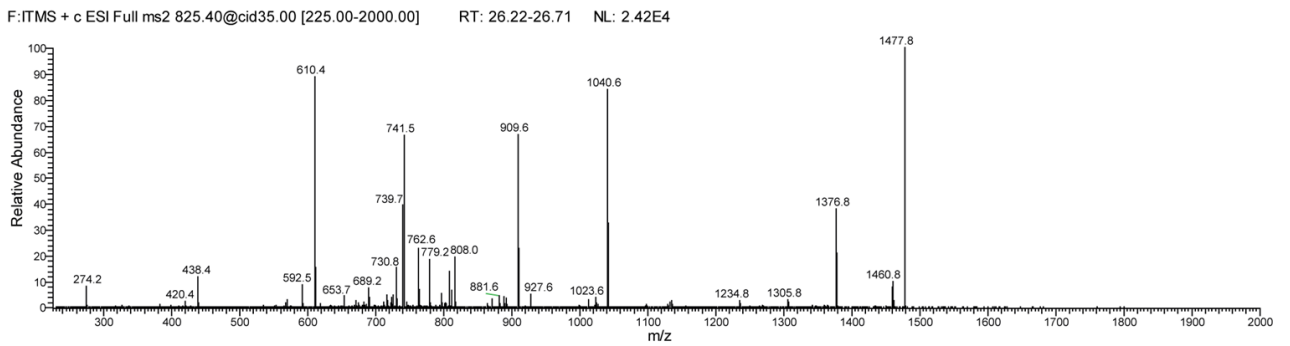


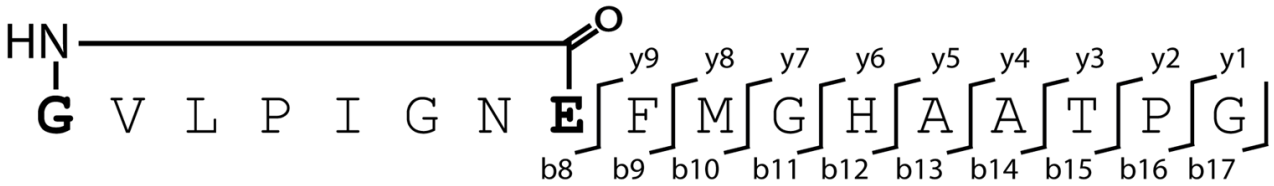


fragment ion	charge	predicted mass	observed mass	intensity	fragment ion	charge	predicted mass	observed mass	intensity
b16	+2	822.4	822.7	1710	**b10-H_2O**	+1	925.4	925.4	3500
b16-H_2O	+2	813.4	814.1	151000	**b9**	+1	830.3	830.5	120
a16	+2	808.4	808.8	95100	**b9-H_2O**	+1	812.3	812.4	14200
a16-H_2O	+2	799.4	800.1	32200	**b8**	+1	715.3	715.4	57900
b15	+2	772.9	773.2	144050	**b8-H_2O**	+1	697.3	697.4	10600
b15-H_2O	+2	763.9	764.4	108000	**y2**	+1	233.1	233.1	375
b15	+1	1544.7	1544.7	10800	**y3**	+1	380.2	380.3	648
b15-H_2O	+1	1526.7	1526.8	2850	**x4-NH**	+1	488.2	488.4	15600
b14	+1	1429.7	1429.6	3620	**y4-H_2O**	+1	459.2	459.3	5770
b13	+1	1282.6	1282.7	459	**y4**	+1	477.2	477.3	35800
b12	+1	1185.6	1185.5	15300	**y5-H_2O**	+1	587.3	587.4	11100
b12-H_2O	+1	1167.5	1167.6	25200	**y5**	+1	605.3	605.4	76900
b12-2H_2O	+1	1149.5	1149.6	3030	**y6**	+1	719.4	719.5	3840
b11	+1	1057.5	1057.4	23900	**y7**	+1	832.5	832.7	44
b11-H_2O	+1	1039.4	1039.4	6640	**y8**	+1	947.5	947.6	4210
b10	+1	943.4	943.4	54800					

Supplementary Figure S6 - Mass analysis of zucinodin. On top the total ion current (TIC), extracted ion current (EIC) for the singly and doubly charged ions of the main product (full length) and the UV signal of the 3 d 20 °C pellet extract are shown. Depicted in the middle is the fragmentation spectrum of the doubly charged main product. All detectable fragment ions with their predicted masses, observed masses and signal intensities are listed on the bottom.

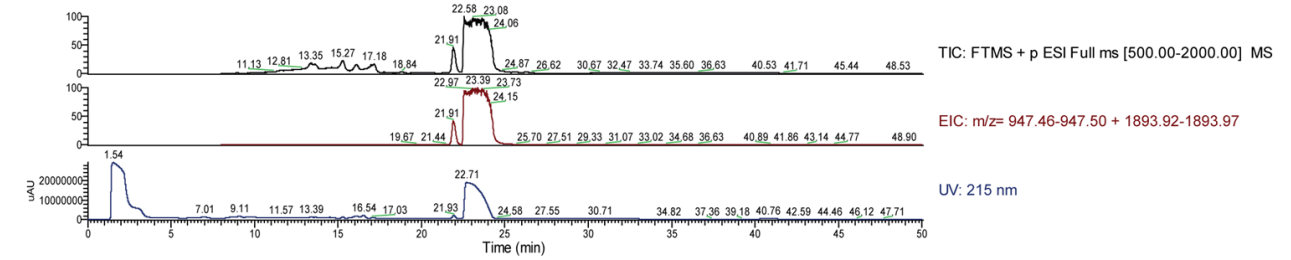


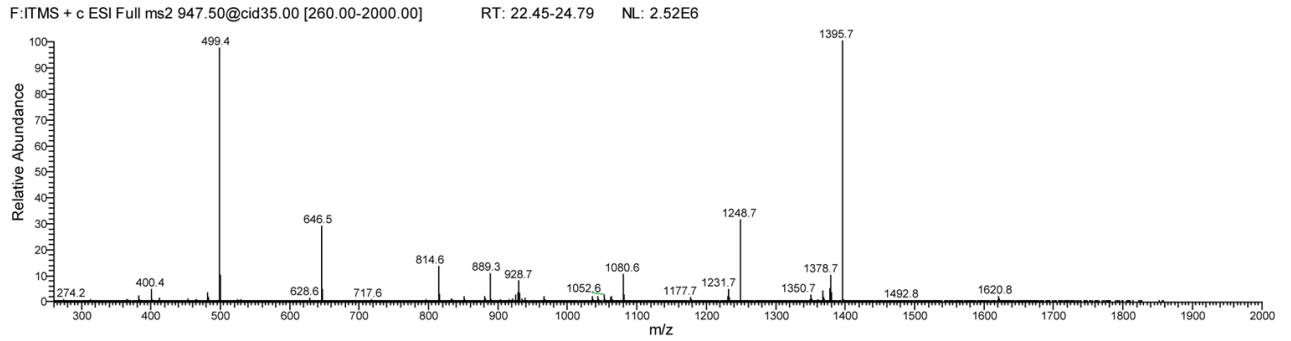


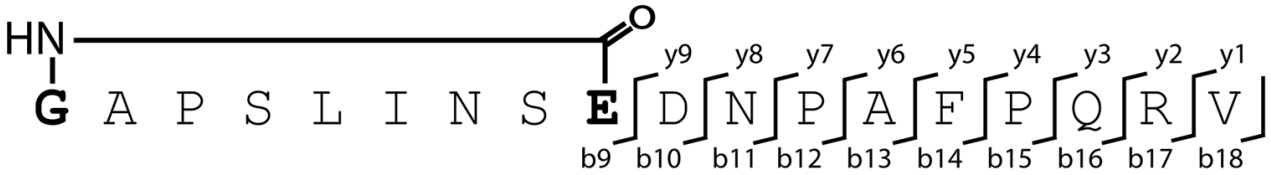


fragment ion	charge	predicted mass	observed mass	intensity	fragment ion	charge	predicted mass	observed mass	intensity
b17	+2	816.4	816.7	4690	**a10**	+1	1012.5	1012.5	669
b17-NH_3	+2	807.9	808.0	3390	**b9**	+1	909.5	909.6	16100
b15	+2	739.4	739.7	9550	**a9**	+1	881.5	881.6	1050
b15-H_2O	+2	730.4	730.8	3760	**c8**	+1	779.3	779.2	4510
b14	+2	688.8	689.2	1780	**b8**	+1	762.4	762.6	5510
b13	+2	653.3	653.7	1090	**y3**	+1	274.2	274.2	1990
b15	+1	1477.7	1477.8	24200	**y4**	+1	345.2	345.3	15
b15-NH_3	+1	1460.7	1460.8	2400	**y5**	+1	416.2	416.4	114
b14	+1	1376.7	1376.8	9190	**y6**	+1	553.3	553.4	195
b13	+1	1305.6	1305.8	685	**y7-H_2O**	+1	592.3	592.5	2090
b12	+1	1234.6	1234.8	636	**y7**	+1	610.3	610.4	21600
b11	+1	1097.6	1097.7	229	**y8**	+1	741.3	741.5	16100
b10	+1	1040.5	1040.6	2040	**y9**	+1	888.4	888.6	1010
b10-NH_3	+1	1023.6	1023.6	914					

Supplementary Figure S7 - Mass analysis of rhodanodin. On top the total ion current (TIC), extracted ion current (EIC) for the singly and doubly charged ions of the main product (-3 aa truncation) and the UV signal of the 3 d 20 °C pellet extract are shown. Depicted in the middle is the fragmentation spectrum of the doubly charged main product. All detectable fragment ions with their predicted masses, observed masses and signal intensities are listed on the bottom.

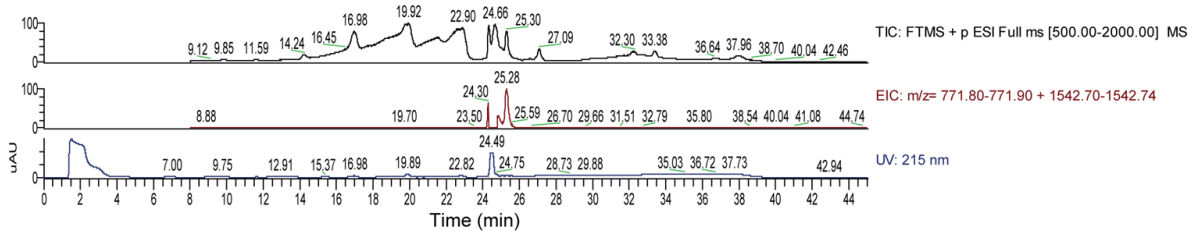


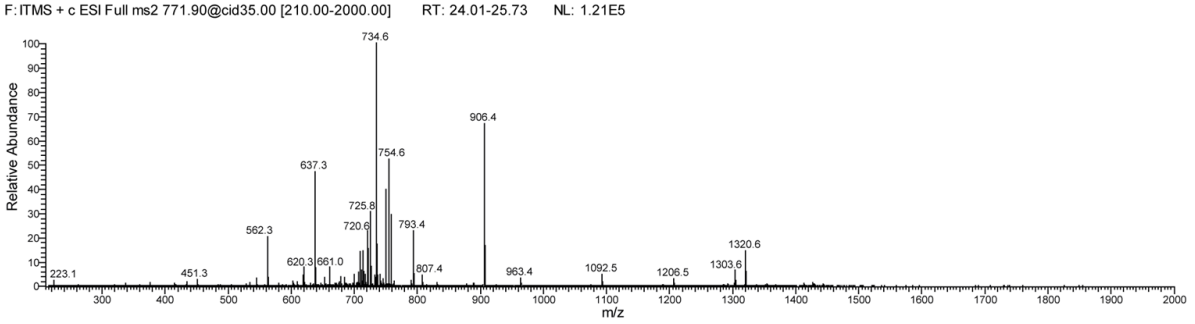


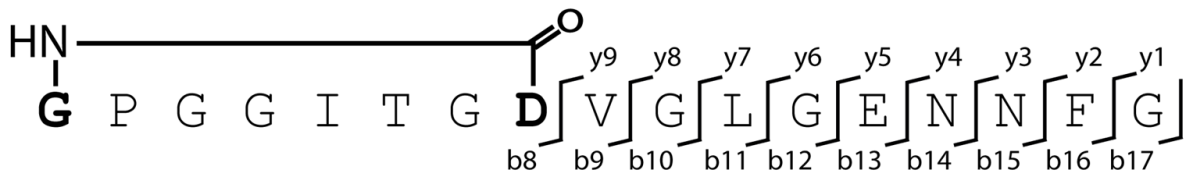


fragment ion	charge	predicted mass	observed mass	intensity	fragment ion	charge	predicted mass	observed mass	intensity
b18-NH_3	+2	930.0	930.0	199000	**a11**	+1	1052.5	1052.6	59300
a18	+2	924.5	924.8	60200	**a11-NH_3**	+1	1052.5	1035.6	39900
b17	+2	888.9	889.3	264000	**b10**	+1	966.5	966.6	43800
b16	+1	1620.8	1620.8	4600	**a10**	+1	938.5	938.8	23000
b16-NH_3	+1	1603.7	1603.8	6110	**b9**	+1	851.4	851.6	39600
b15	+1	1492.7	1492.8	6020	**y2**	+1	274.2	274.2	13000
b14	+1	1395.7	1395.7	2520000	**y4-NH_3**	+1	481.3	481.4	83800
b14-NH_3	+1	1378.6	1378.7	248000	**y4**	+1	499.3	499.4	2450000
a14	+1	1367.7	1367.6	93500	**y5-H_2O**	+1	628.4	628.6	28400
a14-NH_3	+1	1350.7	1350.7	58900	**y5**	+1	646.4	646.5	726000
b13	+1	1248.6	1248.7	790000	**y6**	+1	717.4	717.6	13700
b13-NH_3	+1	1231.6	1231.7	110000	**y7-H_2O**	+1	796.5	796.7	13000
b12	+1	1177.6	1177.7	31200	**y7**	+1	814.5	814.6	336000
b11	+1	1080.5	1080.6	255000	**y8**	+1	928.5	928.7	82300
b11-NH_3	+1	1063.5	1063.7	43000	**y9**	+1	1043.5	1043.7	44700

Supplementary Figure S8 - Mass analysis of rubrivinodin. On top the total ion current (TIC), extracted ion current (EIC) for the singly and doubly charged ions of the main product (full length) and the UV signal of the 1 d 37 °C pellet extract are shown. Depicted in the middle is the fragmentation spectrum of the doubly charged main product. All detectable fragment ions with their predicted masses, observed masses and signal intensities are listed on the bottom.

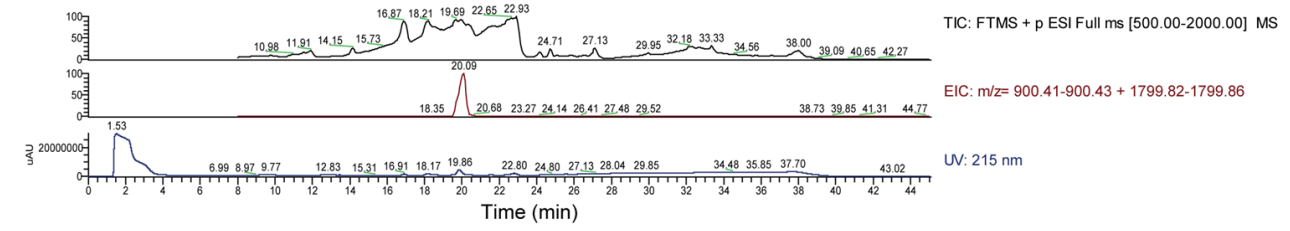


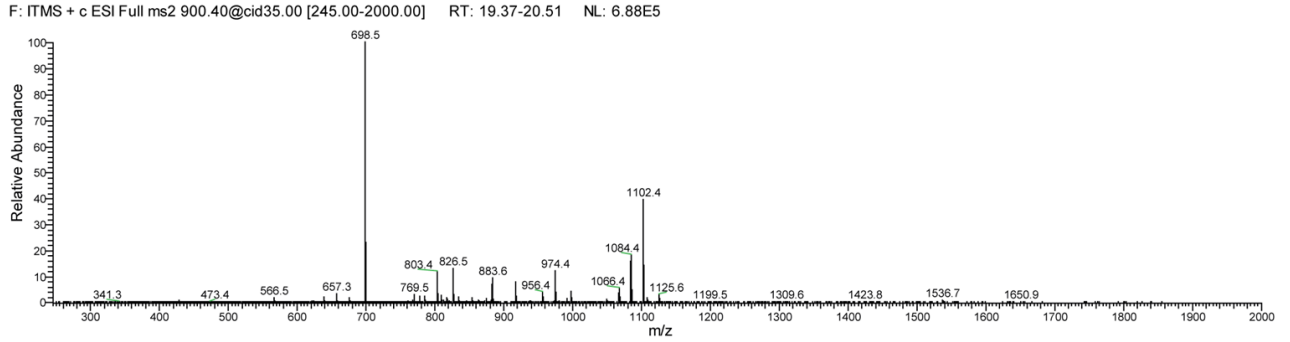


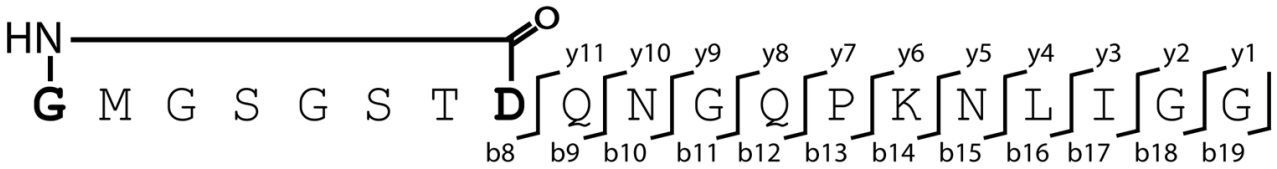


fragment ion	charge	predicted mass	observed mass	intensity	fragment ion	charge	predicted mass	observed mass	intensity
b17	+2	762.9	762.8	2430	**b9**	+1	736.4	736.4	20800
b17-H_2O	+2	753.9	754.6	63200	**a9**	+1	708.4	708.6	17200
b16	+2	734.4	734.6	121000	**c8**	+1	654.3	654.3	702
b16-H_2O	+2	725.3	725.8	36900	**b8**	+1	637.3	637.3	56800
a16	+2	720.4	720.6	27200	**b8-H_2O**	+1	619.3	619.4	5280
a16-H_2O	+2	711.4	711.9	7990	**a8**	+1	609.3	609.4	2120
b15	+2	660.8	661.0	9350	**y2**	+1	223.1	223.1	2710
b15-H_2O	+2	651.8	652.3	4390	**y3**	+1	337.2	337.2	1390
b15	+1	1320.6	1320.6	17300	**y4**	+1	451.2	451.3	3090
b15-NH_3	+1	1303.6	1303.6	7770	**y5-H_2O**	+1	562.2	562.3	24300
b14	+1	1206.6	1206.5	3520	**y5**	+1	580.2	580.3	1220
b14-H_2O	+1	1188.6	1188.6	560	**z6-H_2O**	+1	602.2	602.3	2360
b13	+1	1092.5	1092.4	5600	**y6-H_2O**	+1	619.3	619.4	56800
b13-H_2O	+1	1074.5	1074.5	629	**z6**	+1	620.2	620.3	9230
b12	+1	963.5	963.4	4010	**y6**	+1	637.3	637.3	5280
b12--H_2O	+1	945.5	945.5	331	**y7-2H_2O**	+1	714.3	714.1	17600
b11	+1	906.5	906.4	80700	**y7-H_2O**	+1	732.3	732.4	5420
b11-H_2O	+1	888.5	888.5	1420	**y8**	+1	807.4	807.4	5310
a11	+1	878.5	878.3	994	**y9-H_2O**	+1	906.4	906.4	80700
b10	+1	793.4	793.4	27400	**y9**	+1	888.4	888.5	1420
b10-H_2O	+1	775.4	775.5	135					

Supplementary Figure S9 - Mass analysis of sphingonodin I. On top the total ion current (TIC), extracted ion current (EIC) for the singly and doubly charged ions of the main product (-4 aa truncation) and the UV signal of the 3 d 20 °C pellet extract are shown. Depicted in the middle is the fragmentation spectrum of the doubly charged main product. All detectable fragment ions with their predicted masses, observed masses and signal intensities are listed on the bottom.

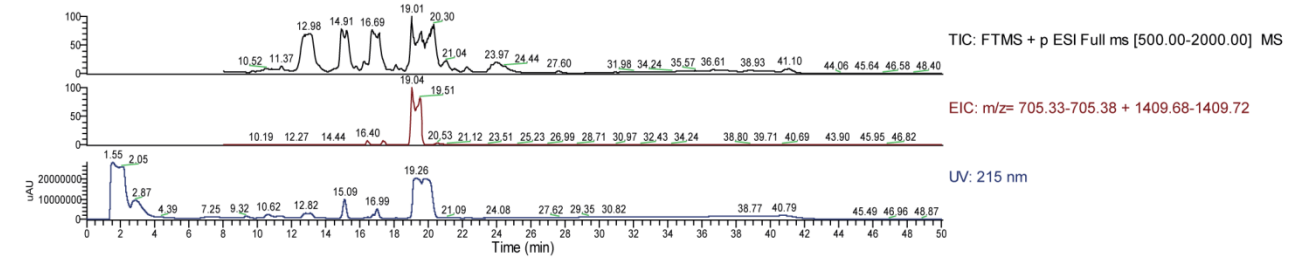


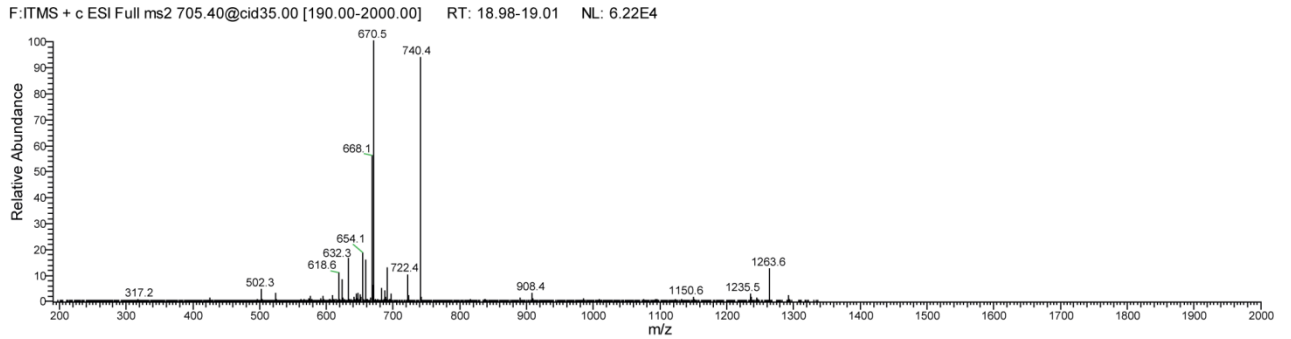


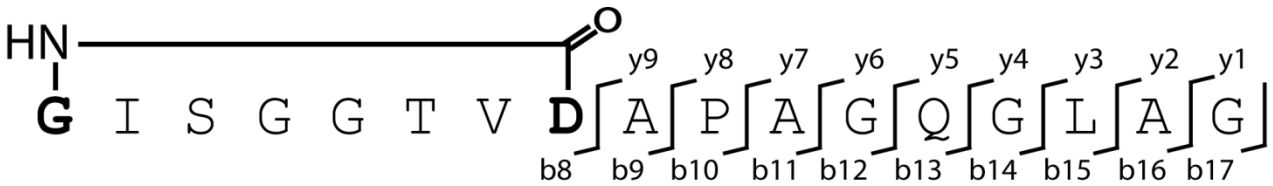


fragment ion	charge	predicted mass	observed mass	intensity	fragment ion	charge	predicted mass	observed mass	intensity	fragment ion	charge	predicted mass	observed mass	intensity
b19	+2	891.4	891.6	337	**b14-H_2O**	+1	1309.6	1309.6	1980	**b8-H_2O**	+1	657.2	657.3	21700
b19-H_2O	+2	882.4	882.8	47100	**b14-$2H_2O$**	+1	1291.6	1291.5	140	**b8-$2H_2O$**	+1	639.2	639.3	13100
b19-$2H_2O$	+2	873.4	874.2	9650	**b13**	+1	1199.5	1199.5	793	**y4-H_2O**	+1	341.2	341.3	659
b18	+2	862.9	863.3	4950	**b13-H_2O**	+1	1181.5	1181.5	385	**y4**	+1	359.2	359.3	327
b18-H_2O	+2	853.9	854.5	12200	**b12**	+1	1102.4	1102.4	272000	**y5**	+1	473.3	473.4	1570
b18-$2H_2O$	+2	844.9	845.7	2330	**b12-H_2O**	+1	1084.4	1084.4	125000	**z6-H_2O**	+1	566.3	566.5	11500
b17	+2	834.4	834.6	13900	**b12-$2H_2O$**	+1	1066.4	1066.4	36900	**y6**	+1	601.4	601.5	1740
b17-H_2O	+2	816.4	817.1	11600	**b12-$3H_2O$**	+1	1048.4	1048.5	7210	**y7**	+1	698.4	698.5	688000
b16	+2	777.9	778.3	15000	**c11**	+1	991.4	991.5	9320	**y8**	+1	826.5	826.5	89700
b16-H_2O	+2	768.9	769.5	19700	**b11**	+1	974.4	974.4	81900	**y9**	+1	883.5	883.6	64400
b16	+1	1554.7	1554.7	1520	**b11-H_2O**	+1	956.4	956.4	25300	**z10**	+1	980.5	980.7	3130
b16-H_2O	+1	1536.7	1536.7	4390	**b11-$2H_2O$**	+1	938.3	938.4	3640	**y10**	+1	997.5	997.6	28800
b16-$2H_2O$	+1	1518.7	1518.5	532	**b10**	+1	917.3	917.4	53000	**y11-H_2O**	+1	1107.6	1107.7	10500
b15	+1	1441.6	1441.7	342	**b9**	+1	803.3	803.4	81300	**y11**	+1	1125.6	1125.6	20300
b15-H_2O	+1	1423.6	1423.8	751	**b9-H_2O**	+1	785.3	785.4	16200	**y11**	+2	563.3	563.5	782
b14	+1	1327.6	1327.6	718	**b8**	+1	675.2	675.3	10700					

Supplementary Figure S10 - Mass analysis of sphingonodin II. On top the total ion current (TIC), extracted ion current (EIC) for the singly and doubly charged ions of the main product (-4 aa truncation) and the UV signal of the 3 d 20 °C pellet extract are shown. Depicted in the middle is the fragmentation spectrum of the doubly charged main product. All detectable fragment ions with their predicted masses, observed masses and signal intensities are listed on the bottom.

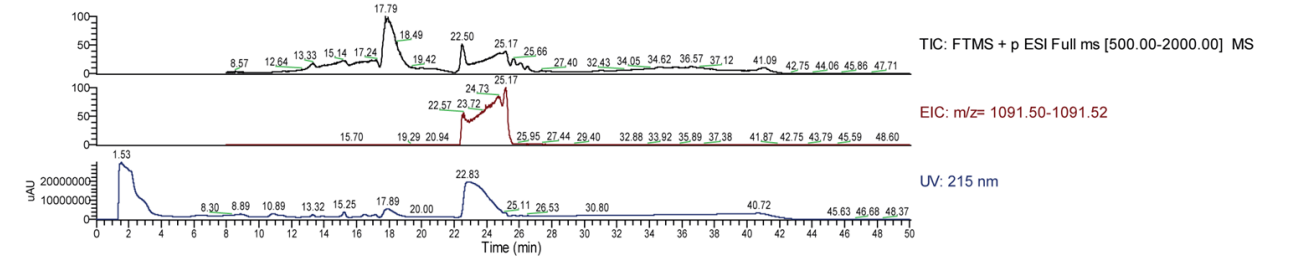


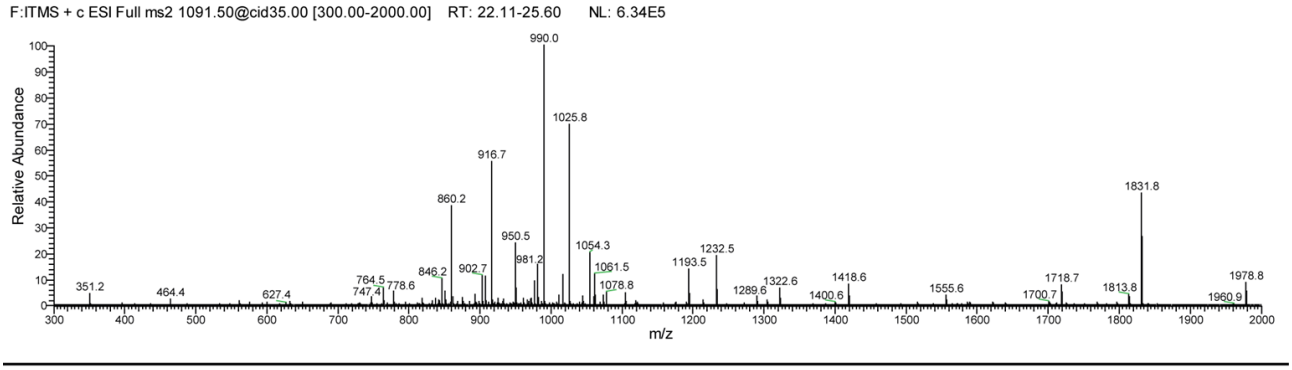


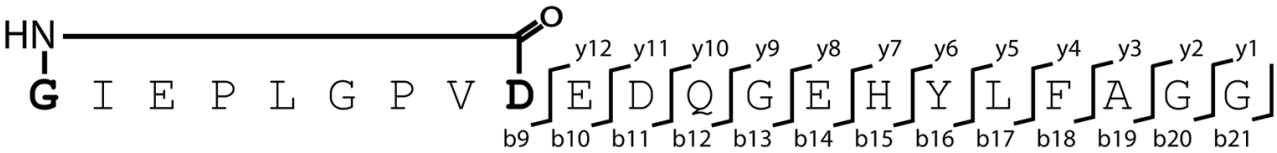


fragment ion	charge	predicted mass	observed mass	intensity	fragment ion	charge	predicted mass	observed mass	intensity
b17	+2	696.4	696.6	1010	**b12**	+1	965.5	965.4	43
b16	+2	667.8	668.1	32200	**b11**	+1	908.5	908.4	1830
b16-H_2O	+2	658.8	659.0	7050	**b10**	+1	837.4	837.4	175
a16	+2	653.8	654.1	9860	**b9**	+1	740.4	740.4	57700
b15	+2	632.3	632.5	9180	**b9-H_2O**	+1	722.4	722.4	6110
b15-H_2O	+2	623.3	623.5	4710	**b8**	+1	669.3	669.3	2970
a15	+2	618.3	618.6	5230	**y3**	+1	260.2	260.1	94
b15	+1	1263.6	1263.6	7810	**y4**	+1	317.2	317.2	316
b15-H_2O	+1	1245.6	1245.6	880	**y5**	+1	445.2	445.6	99
a15	+1	1235.6	1235.5	1410	**y6**	+1	502.3	502.3	2510
b14	+1	1150.6	1150.6	530	**y7**	+1	573.3	573.4	450
b14-H_2O	+1	1132.5	1132.5	181	**y8**	+1	670.4	670.5	62200
b13	+1	1093.5	1093.5	3315	**y9**	+1	741.4	741.4	12000
b13-H_2O	+1	1075.5	1075.5	246					

Supplementary Figure S11 - Mass analysis of syanodin I. On top the total ion current (TIC), extracted ion current (EIC) for the singly and doubly charged ions of the main product (-4 aa truncation) and the UV signal of the 3 d 20 °C pellet extract are shown. Depicted in the middle is the fragmentation spectrum of the doubly charged main product. All detectable fragment ions with their predicted masses, observed masses and signal intensities are listed on the bottom.

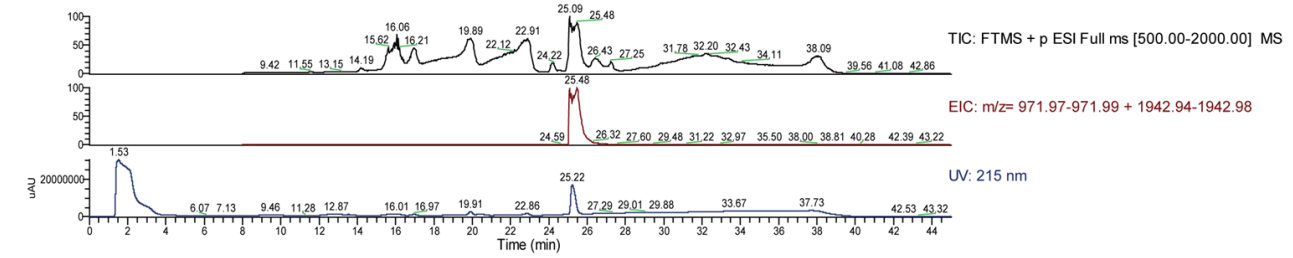


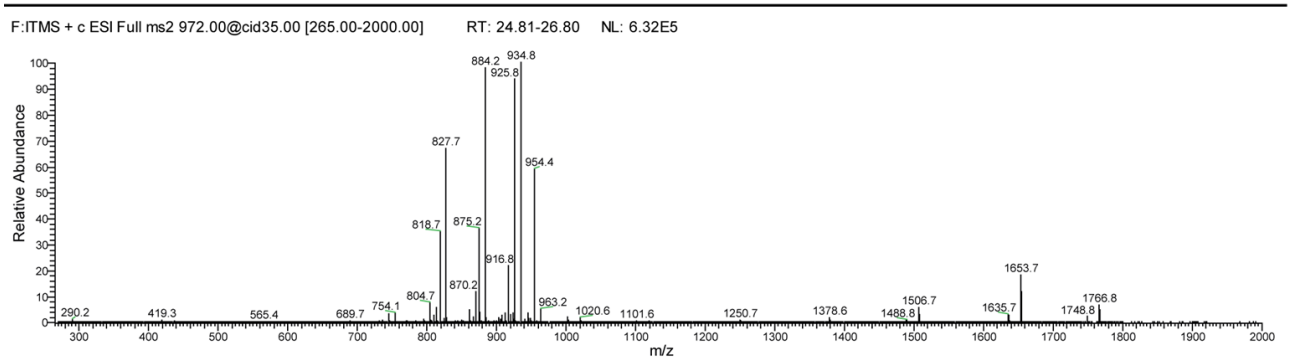


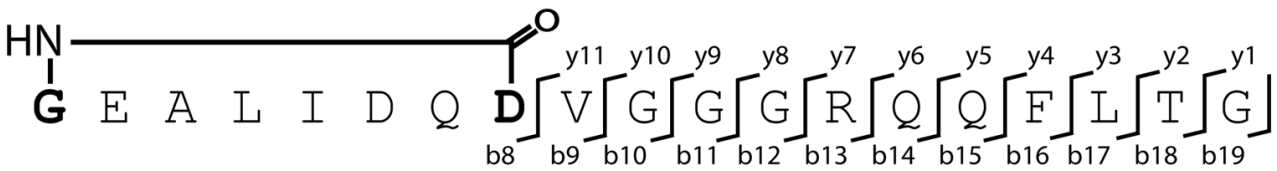


fragment ion	charge	predicted mass	observed mass	intensity	fragment ion	charge	predicted mass	observed mass	intensity	fragment ion	charge	predicted mass	observed mass	intensity
b21	+2	1082.5	1082.6	392	**c15**	+2	786.9	786.7	3380	**b11**	+1	1104.5	1104.5	30000
b21-H_2O	+2	1073.5	1073.8	23900	**b15**	+2	778.4	778.6	33400	**a10**	+1	961.5	961.2	16900
a21	+2	1068.5	1068.8	6880	**b18**	+1	1978.9	1978.8	56100	**b9**	+1	860.5	860.2	242000
b20	+2	1054.0	1054.3	129000	**b18-H_2O**	+1	1960.9	1960.9	3100	**y4**	+1	351.2	351.2	27800
b20-H_2O	+2	1045.0	1045.2	22600	**b17**	+1	1831.9	1831.8	273000	**y5**	+1	464.3	464.4	14900
b19	+2	1025.5	1025.8	443000	**b17-H_2O**	+1	1813.8	1813.8	25200	**y6**	+1	627.3	627.4	5310
b19-H_2O	+2	1016.5	1016.8	73800	**b16**	+1	1718.8	1718.7	49400	**z7**	+1	747.4	747.4	20500
a19	+2	1011.5	1011.8	23900	**b16-H_2O**	+1	1700.8	1700.7	8930	**y7**	+1	764.4	764.5	41900
b18	+2	990.0	990.0	634000	**b15**	+1	1555.7	1555.6	24700	**y8-H_2O**	+1	875.4	875.5	18300
b18-H_2O	+2	981.0	981.2	99900	**b15-H_2O**	+1	1537.7	1537.7	3660	**y8**	+1	893.4	893.6	25400
a18	+2	976.0	976.3	59200	**b14**	+1	1418.6	1418.6	51200	**y9**	+1	950.4	950.5	153000
a18-H_2O	+2	967.0	967.2	12100	**b14-H_2O**	+1	1400.6	1400.6	7680	**y10**	+1	1078.5	1078.8	31000
c17	+2	924.9	925.6	16200	**b13**	+1	1289.6	1289.6	21500	**z11-H_2O**	+1	1158.5	1158.5	4580
b17	+2	916.4	916.7	351000	**b13-H_2O**	+1	1271.6	1271.6	4130	**z11**	+1	1176.5	1176.6	5970
b17-H_2O	+2	907.4	907.7	71000	**b12**	+1	1232.6	1232.5	120000	**y11**	+1	1193.5	1193.5	88100
a17	+2	902.4	902.7	72000	**b12-H_2O**	+1	1214.6	1214.6	12400	**y12-H_2O**	+1	1304.6	1304.6	12000
a16	+2	845.9	846.2	64800	**c11**	+1	1121.6	1121.6	7490	**y12**	+1	1322.6	1322.6	41800

Supplementary Figure S12 - Mass analysis of sphingopyxin I. On top the total ion current (TIC), extracted ion current (EIC) for the doubly charged ions of the main product (-5 aa truncation) and the UV signal of the 3 d 20 °C pellet extract are shown. Depicted in the middle is the fragmentation spectrum of the doubly charged main product. All detectable fragment ions with their predicted masses, observed masses and signal intensities are listed on the bottom.

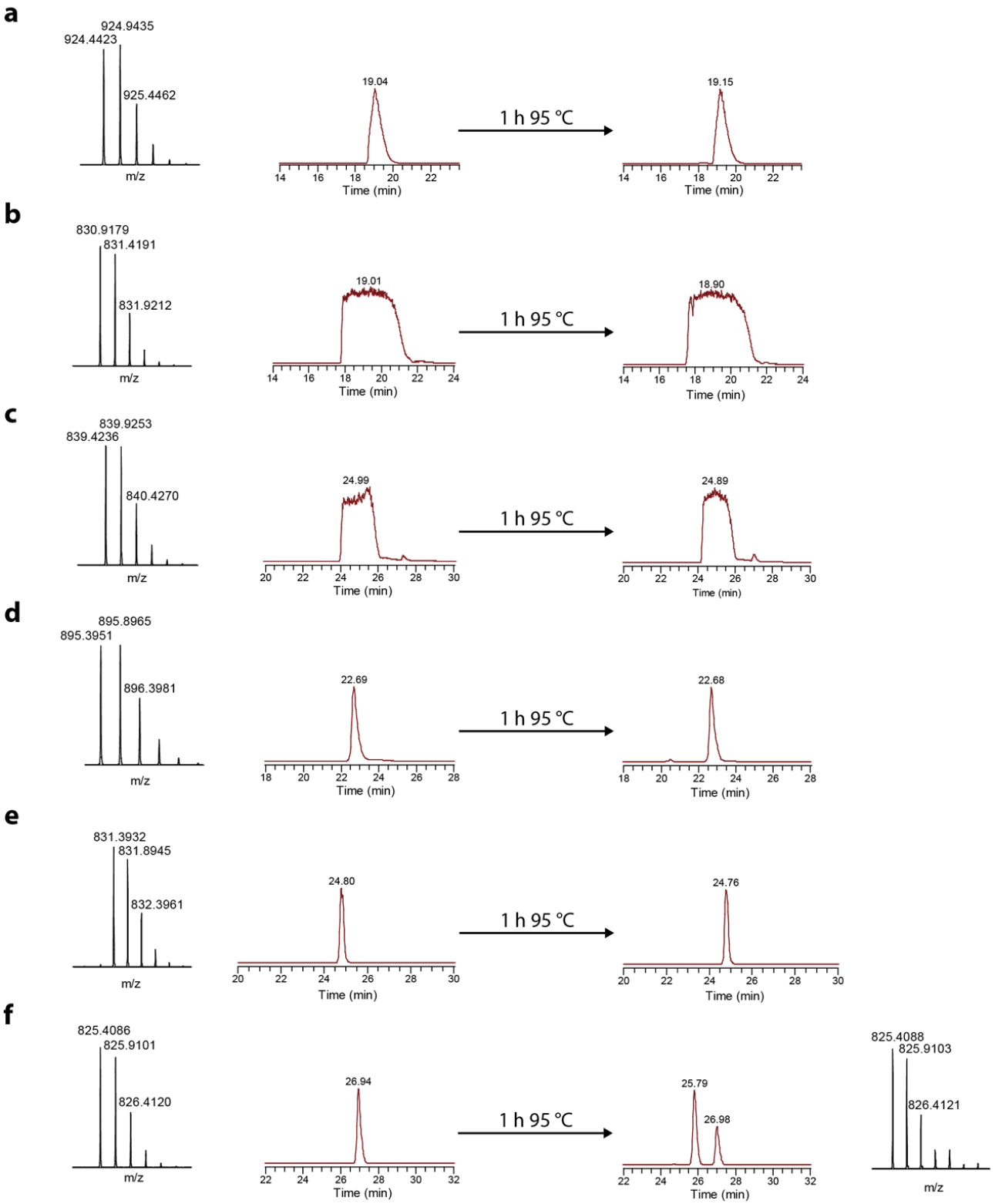




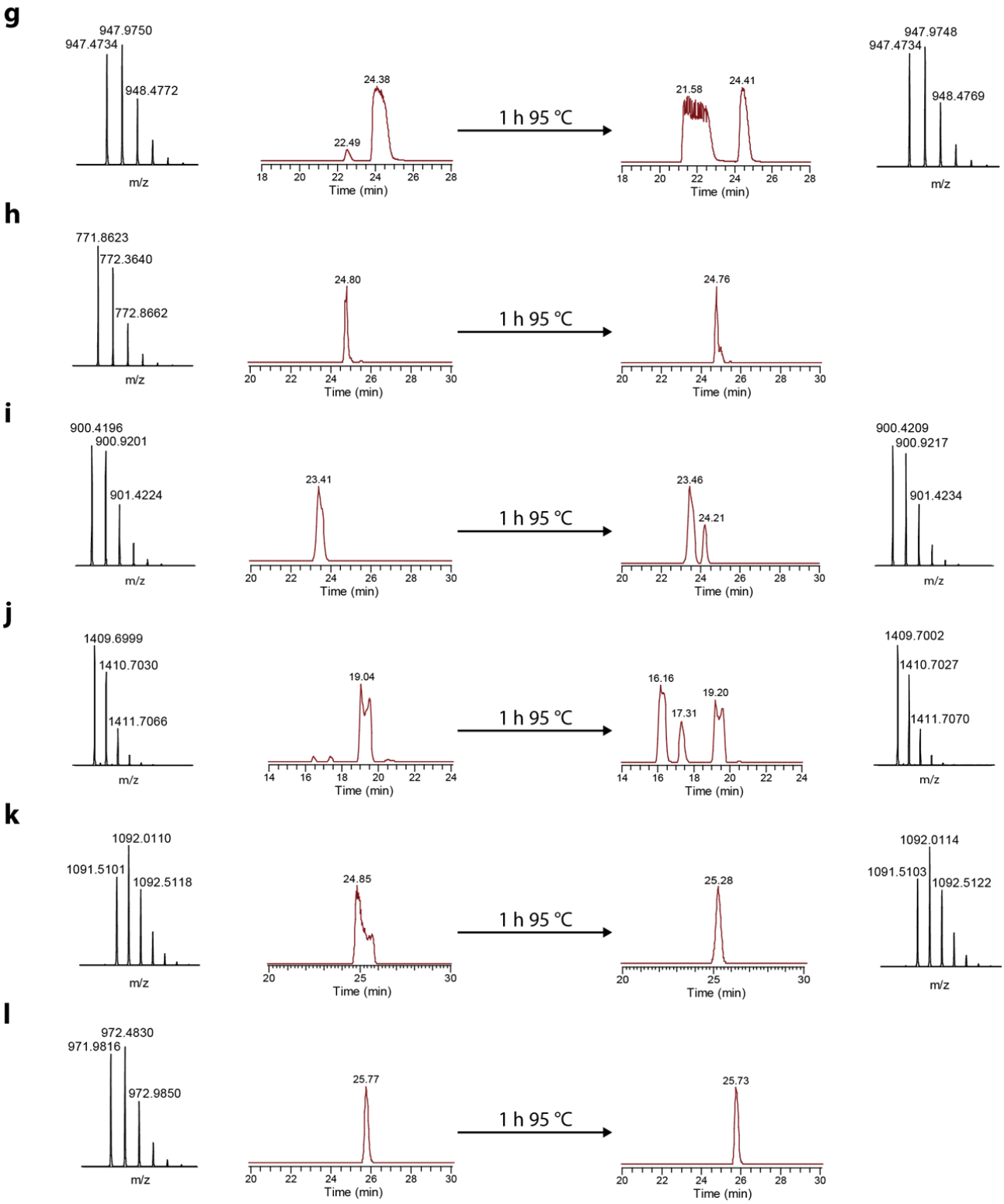


fragment ion	charge	predicted mass	observed mass	intensity	fragment ion	charge	predicted mass	observed mass	intensity	fragment ion	charge	predicted mass	observed mass	intensity
b19-H_2O	+2	954.0	954.4	373000	**b15-2H_2O**	+2	735.9	736.5	3940	**b9**	+1	923.5	923.5	22400
b19-2H_2O	+2	945.0	945.5	22900	**b14**	+2	698.4	689.7	2210	**b8**	+1	824.4	824.4	8550
b18	+2	934.5	934.8	632000	**b17**	+1	1766.9	1766.8	42000	**y3-H_2O**	+1	272.2	272.2	1550
b18-H_2O	+2	925.5	925.8	613000	**b17-H_2O**	+1	1748.9	1748.8	1750	**y3**	+1	290.2	290.2	5930
b18-2H_2O	+2	916.5	916.8	145000	**b16**	+1	1653.8	1653.7	118000	**y4-H_2O**	+1	419.2	419.3	5760
b18-3H_2O	+2	907.5	908.2	17300	**b16-H_2O**	+1	1635.8	1635.7	19500	**y4**	+1	437.2	437.3	2910
b17	+2	884.0	884.2	640000	**b15**	+1	1506.7	1506.7	36600	**y5-H_2O**	+1	547.3	547.5	1350
b17-H_2O	+2	874.9	875.2	245000	**b15-H_2O**	+1	1488.7	1488.8	5430	**y5**	+1	565.3	565.4	1550
a17	+2	870.0	870.2	76000	**b15-2H_2O**	+1	1470.7	1470.7	251	**y6-2H_2O**	+1	657.3	657.4	1050
a17-H_2O	+2	861.0	861.3	30500	**b14**	+1	1378.7	1378.6	10300	**y6-H_2O**	+1	675.4	675.5	1150
b16	+2	827.4	827.7	437000	**b14-H_2O**	+1	1360.7	1360.7	1410	**y6**	+1	693.4	693.4	768
b16-H_2O	+2	818.4	818.7	229000	**b13**	+1	1250.6	1250.7	5260	**y7**	+1	849.5	849.5	3980
a16	+2	813.4	813.7	36700	**b13-H_2O**	+1	1232.6	1232.6	1170	**y8**	+1	906.5	906.2	6600
b16-2H_2O	+2	809.4	809.9	16600	**b12**	+1	1094.5	1094.4	1830	**y9**	+1	963.5	963.2	32300
a16-H_2O	+2	804.4	804.7	48100	**b12-H_2O**	+1	1076.5	1076.5	165	**y10-H_2O**	+1	1002.5	1002.6	12600
c15	+2	762.4	762.1	1720	**b11**	+1	1037.5	1037.6	895	**y10**	+1	1020.5	1020.6	11000
b15	+2	753.9	754.1	22400	**b11-H_2O**	+1	1019.5	1019.5	236	**y11-H_2O**	+1	1101.6	1101.6	2950
b15-H_2O	+2	744.9	745.2	19600	**b10**	+1	980.5	980.5	451	**y11**	+1	1119.6	1119.7	2130

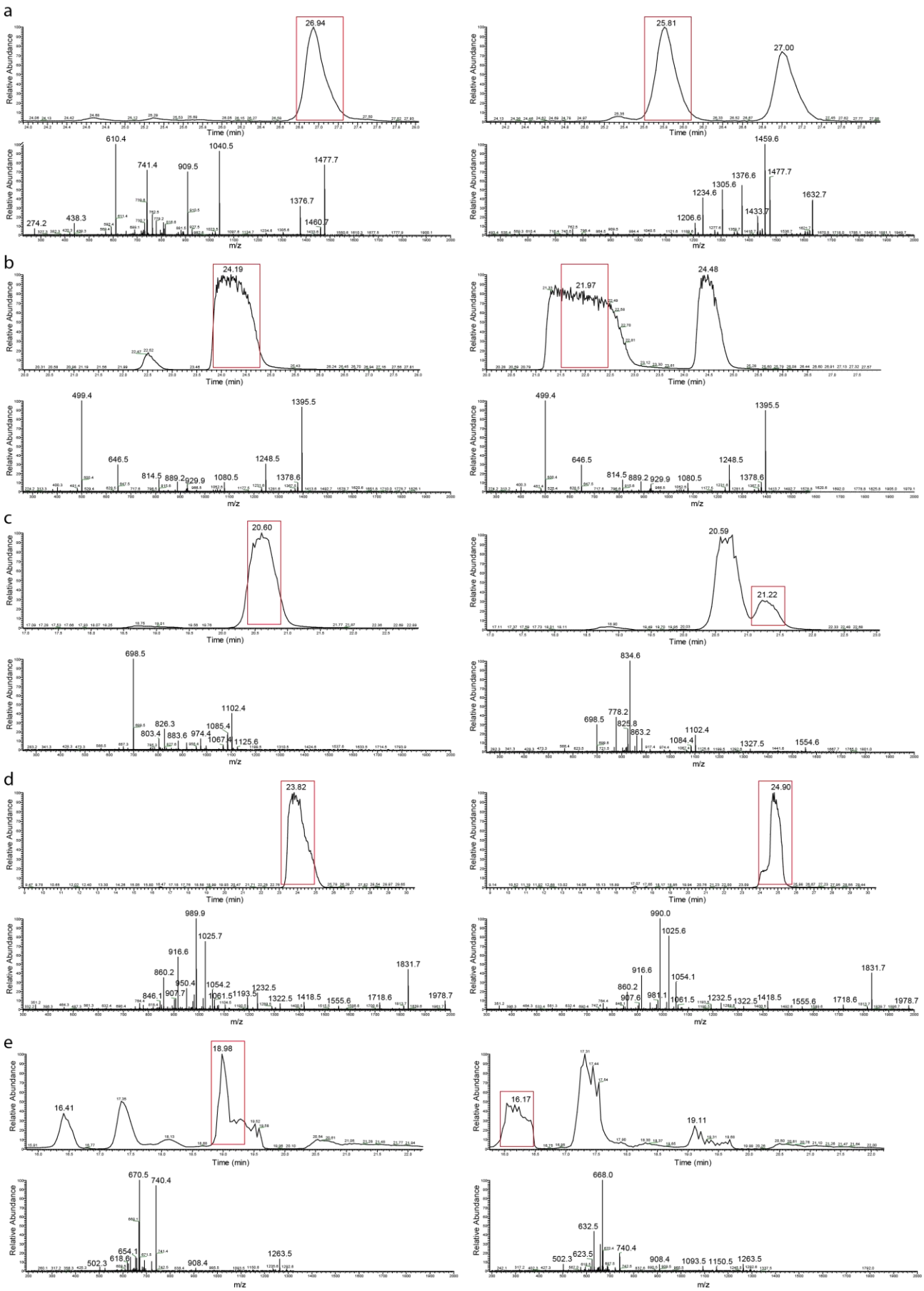
Supplementary Figure S13 - Mass analysis of sphingopyxin II. On top the total ion current (TIC), extracted ion current (EIC) for the doubly charged ions of the main product (-4 aa truncation) and the UV signal of the 3 d 20 °C pellet extract are shown. Depicted in the middle is the fragmentation spectrum of the doubly charged main product. All detectable fragment ions with their predicted masses, observed masses and signal intensities are listed on the bottom.



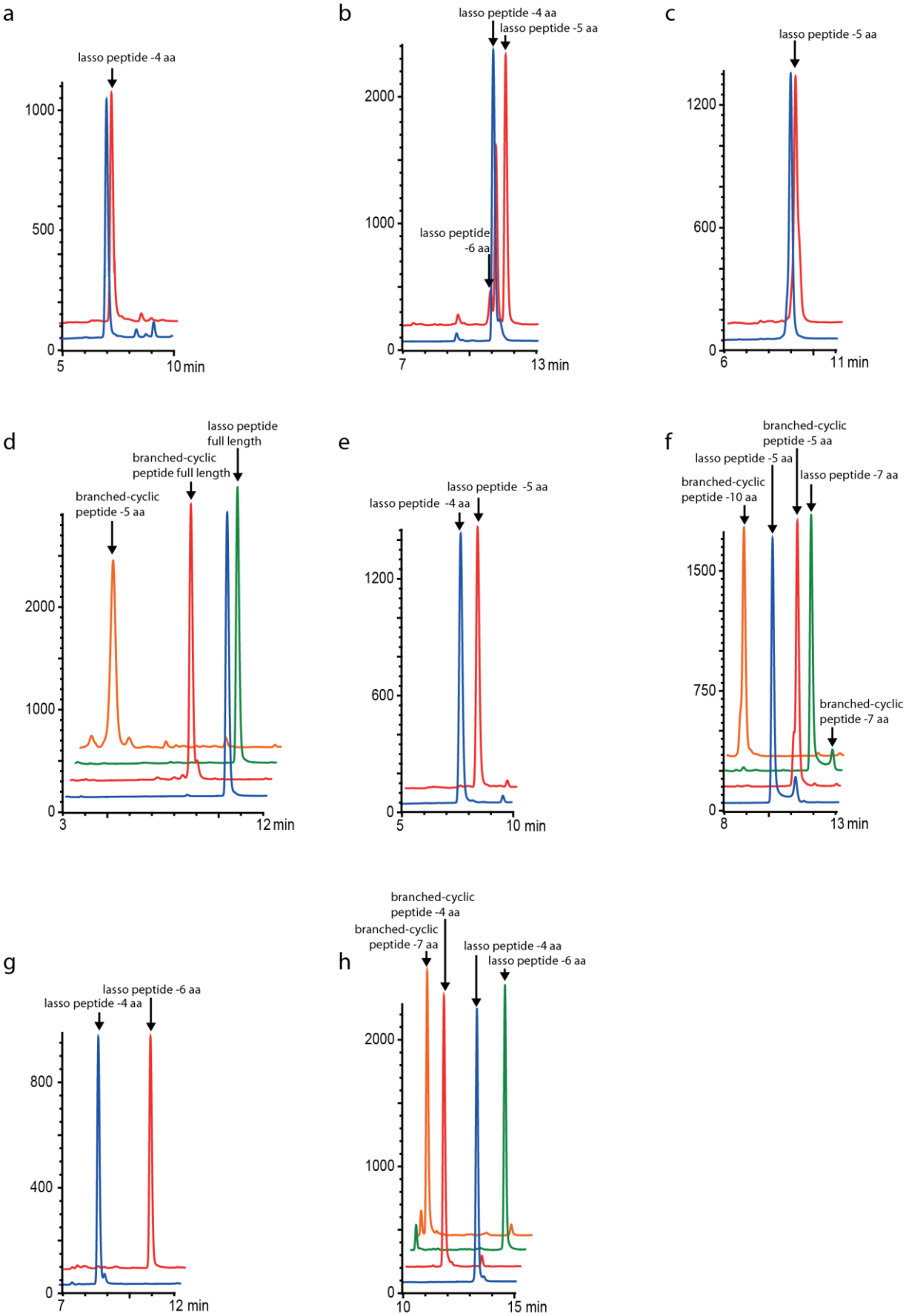
Supplementary Figure S14 -Thermal stability assays for (a) burhizin, (b) caulonodin I, (c) caulonodin II, (d) caulonodin III, (e) zucinodin and (f) rhodanodin. The observed masses and an excerpt of the EIC of the respective lasso peptides are shown on the left. An excerpt of the EIC after incubation for 1 h at 95 °C is presented on the right. In cases where the formation of a second peak was detected, the observed mass of the new compound is shown on the far right. Theoretical m/z values are 924.4433 for burhizin, 830.9186 for caulonodin I, 839.4250 for caulonodin II, 895.3966 for caulonodin III, 831.9444 for zucinodin and 825.4093 for rhodanodin.



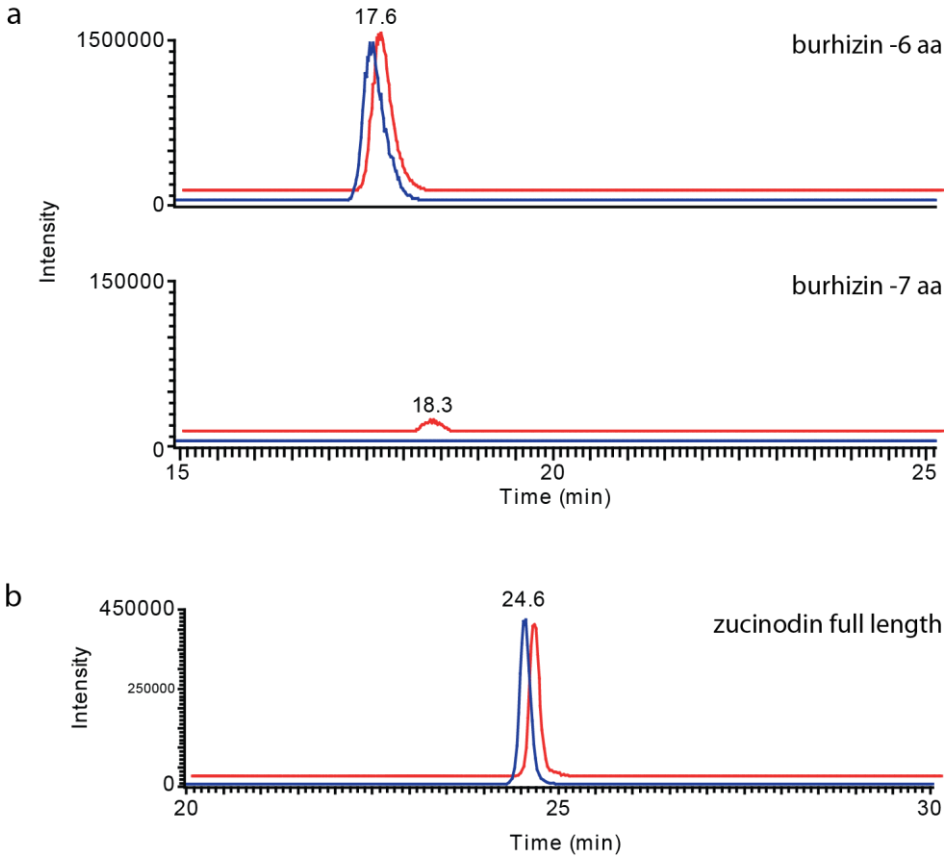
Supplementary Figure S14 - Thermal stability assays for (g) rubrivinodin, (h) sphingonodin I, (i) sphingonodin II, (j) syanodin, (k) sphingopyxin I and (l) sphingopyxin II. The observed masses and an excerpt of the EIC of the respective lasso peptides are shown on the left. An excerpt of the EIC after incubation for 1 h at 95 °C is presented on the right. In cases where the formation of a second peak was detected, the observed mass of the new compound is shown on the far right. Theoretical m/z values are 947.4769 for rubrivinodin, 771.8633 for sphingonodin I, 900.4212 for sphingonodin II, 1409.7024 for syanodin, 1091.5085 for sphingopyxin I and 971.9850 for sphingopyxin II.



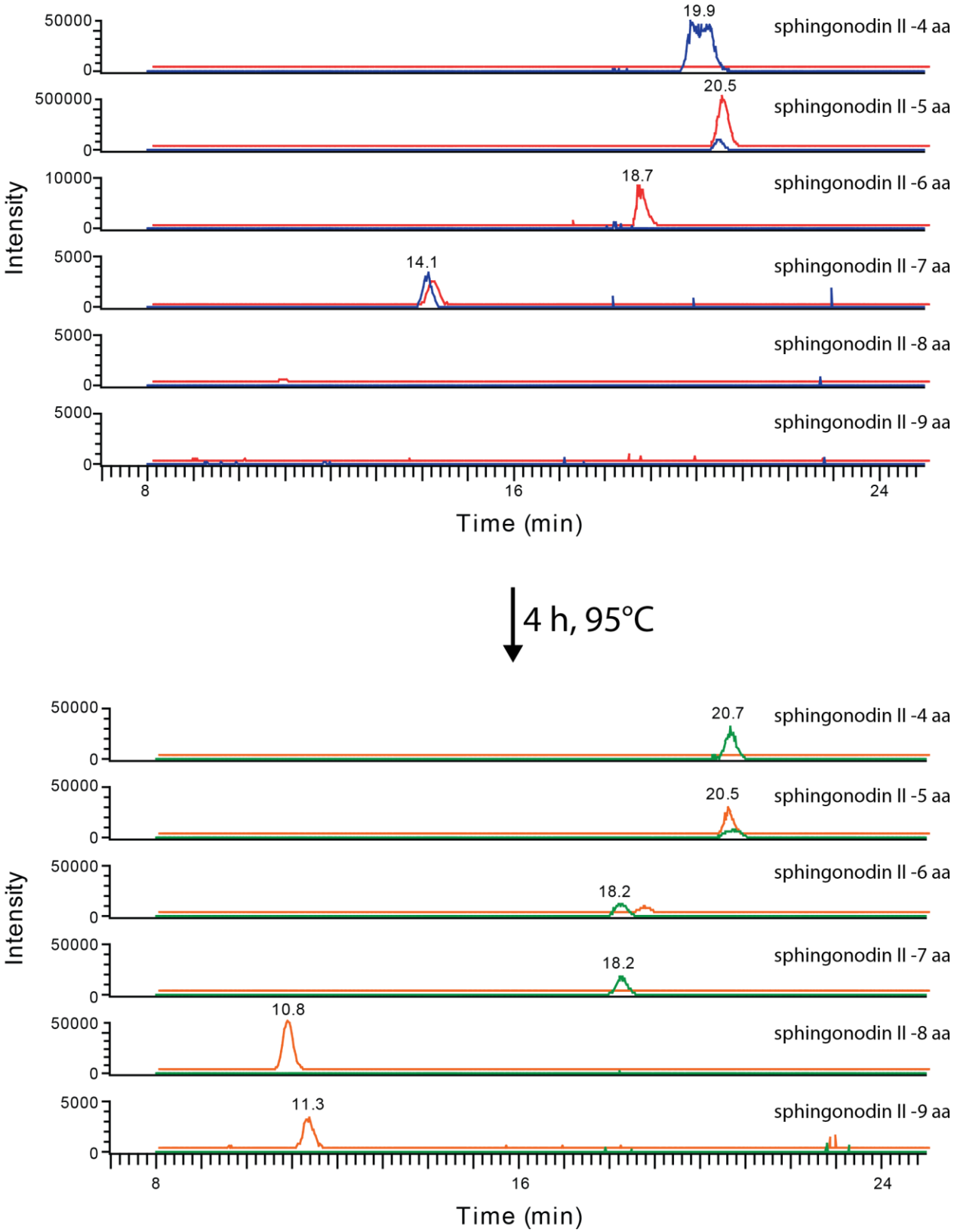
Supplementary Figure S15 – Comparison of the fragmentation spectra of the heat labile lasso peptides (left) and their branched cyclic analogs obtained by incubation for 1 h at 95 °C (right). (a) rhodanodin, (b) rubrivinodin, (c) sphingonodin II, (d) sphingopyxin I and (e) syanodin I.



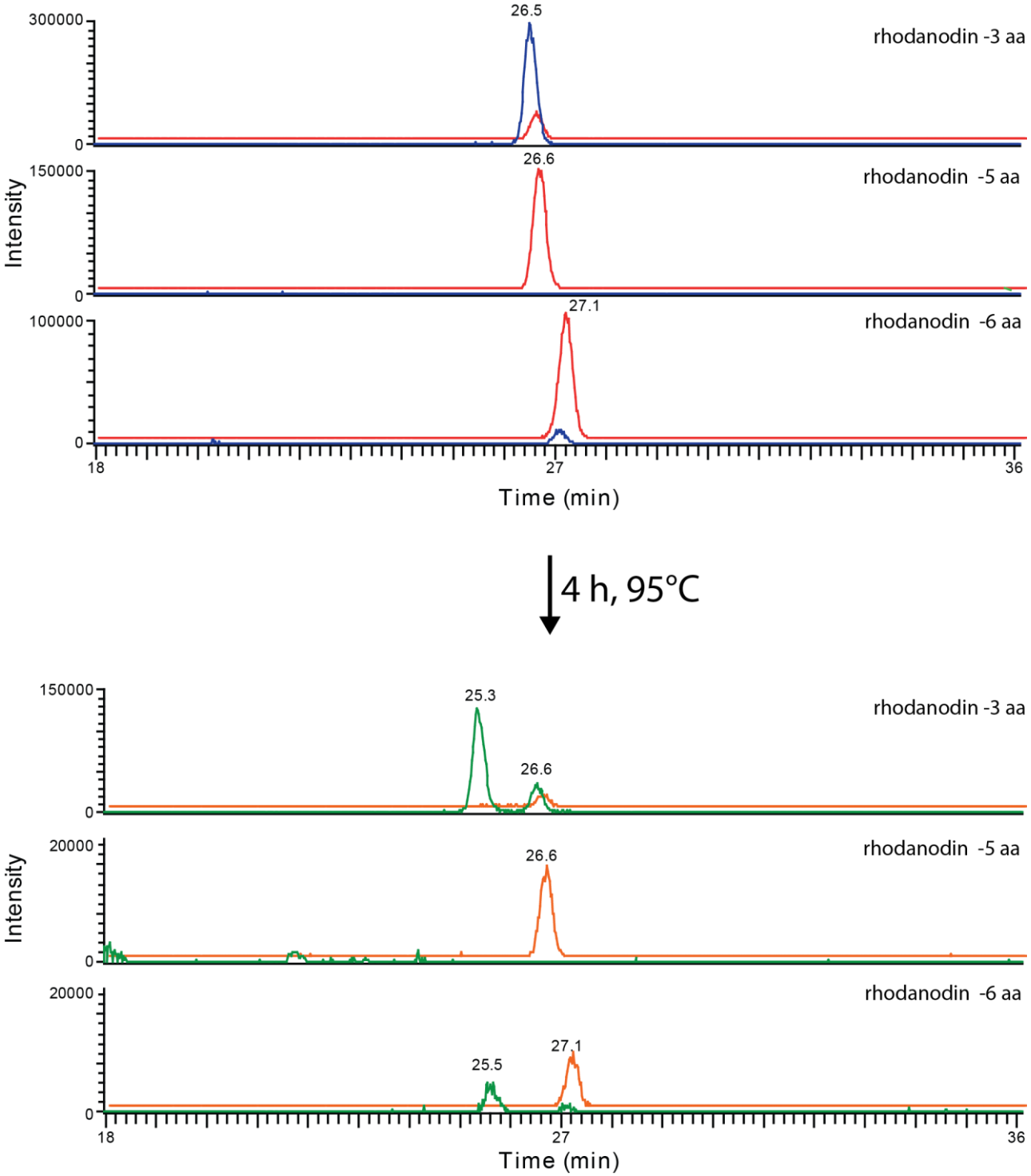
Supplementary Figure S16 – Results of the carboxypeptidase Y digestion assays of the eight isolated lasso peptides (a) caulonodin I, (b) caulonodin II, (c) caulonodin III, (d) rubrivinodin, (e) sphingonodin I, (f) sphingopyxin I, (g) sphingopyxin II and (h) syanodin I. Shown in blue are measurements of the untreated lasso peptides, in red of the lasso peptides after treatment with carboxypeptidase Y for 4 h at 25 °C, in green of the lasso peptides after incubation for 4 h at 95 °C and in orange of the lasso peptides after incubation for 4 h at 95 °C and subsequent carboxypeptidase Y treatment for 4 h at 25 °C.



Supplementary Figure S17 – Results of the carboxypeptidase Y digestion assays with the extracts containing (a) burhizin and (b) zucinodin. The corresponding EIC chromatograms are shown before (blue) and after (red) treatment with carboxypeptidase Y for 4 h at 25 °C. Scales were chosen depending on the signal intensity of the corresponding EICs.



Supplementary Figure S18 – Results of the carboxypeptidase Y digestion assays with the extract containing sphingonodin II. Above, the corresponding EIC chromatograms of unheated extract are shown before (blue) and after (red) treatment with carboxypeptidase Y for 4 h at 25 °C. Below, the corresponding EIC chromatograms of extract after incubation for 4 h at 95 °C are shown before (green) and after (orange) treatment with carboxypeptidase Y for 4 h at 25 °C. Scales were chosen depending on the signal intensity of the corresponding EICs.



Supplementary Figure S19 – Results of the carboxypeptidase Y digestion assays with the extract containing rhodanodin. Above, the corresponding EIC chromatograms of unheated extract are shown before (blue) and after (red) treatment with carboxypeptidase Y for 4 h at 25 °C. Below, the corresponding EIC chromatograms of extract after incubation for 4 h at 95 °C are shown before (green) and after (orange) treatment with carboxypeptidase Y for 4 h at 25 °C. Scales were chosen depending on the signal intensity of the corresponding EICs.

Angewandte Chemie
Eine Zeitschrift der Gesellschaft Deutscher Chemiker

Supporting Information

69451 Weinheim, Germany

Xanthomonins I–III: A New Class of Lasso Peptides with a Seven-Residue Macrolactam Ring**

*Julian D. Hegemann, Marcel Zimmermann, Shaozhou Zhu, Holger Steuber, Klaus Harms, Xiulan Xie, and Mohamed A. Marahiel**

anie_201309267_sm_miscellaneous_information.pdf

Supporting Information

Contents

Experimental Procedures

Bacterial Strains and General Methods. *Xanthomonas gardneri* (DSM no. 19127, ATCC no. 19865) was purchased from the German Collection of Microorganisms and Cell Cultures (DSMZ), while *Xanthomonas citri pv. mangiferaeindicae* (LMG no. 941, ATCC no. 11637) was purchased from the Belgian Co-Ordinated Collections of Micro-Organisms (BCCM/LMG). For cloning and mutagenesis *Escherichia coli* TOP10 was used, while *E. coli* BL21(DE3) was used for heterologous expression. Confirmation of the identity of the constructed plasmids and corresponding mutants was done via DNA dideoxy sequencing by GATC Biotech. Oligonucleotides, proteinase K and carboxypeptidase Y were purchased from Sigma Aldrich, while chymotrypsin was purchased from Promega. Restriction enzymes, Phusion polymerase and T4 DNA Ligase were ordered from New England Biolabs.

Cloning, Modification and Mutagenesis of the Lasso Peptide Biosynthetic Gene Clusters from *Xanthomonas gardneri* and *Xanthomonas citri pv. mangiferaeindicae*. For the cloning of the gene clusters from *X. gardneri* and *X. citri pv. mangiferaeindicae* their DNA was amplified by PCR using Phusion polymerase and appropriate oligonucleotide primers (see Supplementary Table S1a). The ~2.9 kb PCR product from *X. gardneri* was digested with NdeI and SpeI, while the ~2.6 kb PCR product from *X. citri* was digested with NdeI and HindIII. Both clusters were cloned into a likewise digested pET41a vector.

The resulting x*gaA1A2BC* pET41a construct was then modified by PCR using 5'-phosphorylated oligonucleotide primers (see Supplementary Table S1b) flanking the leftovers of the multiple cloning site (MCS). After the PCR reaction, the DNA was purified and subsequently treated with DpnI to remove residual template DNA. Then a blunt end ligation overnight at 16 °C was performed, yielding a variant of the plasmid lacking the remaining MCS. This construct was used for replacing the intergenic region between the *xgaA1A2* and *xgaBC* genes with a terminator sequence followed by the mcjBCD-promoter. This was done via ligation of a PCR amplicon (see Supplementary Table S1b) for the corresponding primers) of the plasmid and an amplicon of the terminator-promoter (T-P) sequence (generated using *csegA1*_T-P_*csegBC* pET41a as template[1]) after digestion with EcoRI and KpnI. Single precursor variants of this construct were created via deletion of either one of the precursor genes by PCR using 5'-phosphorylated oligonucleotide primers (see Supplementary Table S1b) that were flanking the gene to be deleted. The resulting amplicons were isolated and purified, treated with DpnI and finally ligated to yield the single precursor constructs *xgaA1*_T-P_*xgaBC* pET41a and *xgaA2*_T-P_*xgaBC* pET41a. (See Supplementary Table S3a).

The *xcitABC* pET41a plasmid was also modified by PCR using 5'-phosphorylated oligonucleotide primers (see Supplementary Table S1b) to replace the intergenic region between *xcitA* and *xcitBC* with an *E. coli*

optimized ribosomal binding site (RBS). As the forward primer was carrying the RBS as a non-binding overhang on its 5'-end, a blunt end ligation, after DNA purification and a subsequent DpnI digestion, resulted in the desired *xcitA*_RBS_*xcitBC* pET41a construct (See Supplementary Table S3b).

Site-directed ligase-independent mutagenesis (SLIM)[2-3] was used for the generation of xanthomonin II variants. Different sets of SLIM primers were used for introducing mutations either between positions -2 to 9 (see Supplementary Table S1c) or between positions 10 to 20 (see Supplementary Table S1d-S1e) of the precursor peptide. In a similar manner xanthomonin III variants were generated (see Supplementary Table S1f).

In all cases, the ligated or hybridized DNA was dialyzed and then *E. coli* TOP10 was transformed with the DNA via electroporation. After overnight incubation on LB agar plates containing 50 μg/mL kanamycin, individual transformants were used to inoculate liquid LB overnight cultures for plasmid preparations. Isolated plasmids were analyzed by DNA sequencing and, if sequencing confirmed the correct sequences, *E. coli* BL21(DE3) was transformed with the correct constructs for expression.

Production, Isolation and Purification of Xanthomonins I-III. For large scale expressions, M9 minimal medium (17.1 g/L $Na_2HPO_4 \cdot 12\ H_2O$, 3 g/L KH_2PO_4, 0.5 g/L NaCl, 1 g/L NH_4Cl, 1 mL/L $MgSO_4$ solution (2 M), 0.2 mL/L $CaCl_2$ solution (0.5 M), pH = 7.0, after autoclaving 10 mL/L glucose solution (40% w/v) and 0.2 mL/L vitamin mix (see Supplementary Table S2) were added) was inoculated to an OD_{600} of 0.01 with a LB overnight culture of *E. coli* BL21(DE3) carrying the corresponding construct. All cultures contained 50 μg/mL kanamycin.

After inoculation, the cells were grown at 37 °C until an OD_{600} of ~0.4 and then slowly cooled to 20 °C in the course of 1 h. Upon reaching an OD_{600} of ~0.6 the expression was induced by addition of IPTG to a final concentration of 0.05 mM. After 3 days of cultivation at 20 °C, cells were harvested by centrifugation and the supernatant was discarded. For extraction, the cell pellet was resuspended in pure methanol and shaken overnight at 4 °C. For each 1 L of culture fermented, 75 mL of pure methanol was used for extraction. The extract was centrifuged and the supernatant was evaporated at 40 °C and reduced pressure. Dried extract was resuspended in 50% methanol, cleared by centrifugation and filtration and subsequently applied to preparative HPLC (microbore 1100 HPLC system (Agilent) with a VP 250/21 Nucleodur C18 Htec 5 μm column (Macherey-Nagel) at room temperature and a flow rate of 18 mL/min). For xanthomonin I production, the construct *xgaA1*_T-P_*xgaBC* pET41a was expressed in *E. coli* BL21(DE3) and the lasso peptide was purified by two rounds of HPLC. For the first round, the following gradient of water/0.045% formic acid (solvent A) and MeOH/0.05% formic acid (solvent B) was used: Linear increase from 40% to 85% B in 30 min, followed by a linear increase from 85% to 95% B in 5 min and holding 95% B for an additional 3 min. The pooled fractions of the main product (-4 aa truncation of

the full length lasso peptide) were evaporated at 40 °C and reduced pressure and ~2 mg of the semipure compound were subsequently applied to a second round of purification using the following gradient of water/0.1% trifluoroacetic acid (solvent C) and MeCN/0.1% trifluoroacetic acid (solvent D): Linear increase from 20% to 50% D in 30 min, followed by a linear increase from 50% to 95% D in 5 min and holding 95% D for additional 2 min. The pure -4 aa truncation of xanthomonin I was obtained with a yield of 12.7 mg/L.

For xanthomonin II production the construct *xgaA2*_T-P_*xgaBC* pET41a was expressed in *E. coli* BL21(DE3) and purified by two rounds of HPLC. For the first round, the following gradient was used: Linear increase from 25% to 85% B in 60 min, followed by a linear increase from 85% to 95% B in 5 min and holding 95% B for additional 3 min. The pooled fractions of the main product (-6 aa truncation of the full length lasso peptide) were evaporated at 40 °C and reduced pressure and ~2 mg of the semipure compound were applied to a second round of purification using the following gradient: Linear increase from 20% to 50% D in 30 min, followed by a linear increase from 50% to 95% D in 5 min and holding 95% D for additional 2 min. The pure -6 aa truncation of xanthomonin II was obtained with a yield of 4.1 mg/L.

For xanthomonin III production the construct *xcitA*_RBS_*xgaBC* pET41a was expressed in *E. coli* BL21(DE3) and purified by two rounds of HPLC. For the first round, the following gradient was used: Linear increase from 20% to 30% B in 5 min, followed by a linear increase from 30% to 95% B in 25 min and holding 95% B for additional 3 min. The fractions containing the different main products (-4, -5 and -7 aa truncations of the full length lasso peptide) were pooled separately and evaporated at 40 °C and reduced pressure. Subsequently, ~2 mg of the semipure -7 aa truncation were applied to a second round of purification using the following gradient: Linear increase from 10% to 35% D in 30 min, followed by a linear increase from 35% to 95% D in 3 min and holding 95% D for additional 2 min. The pure -7 aa truncation of xanthomonin III was obtained with a yield of 0.3 mg/L.

Expression of Xanthomonin Variants. Xanthomonin II variants were expressed for 3 days at 20 °C in M9 minimal medium as described above. The wild type and each variant were fermented as triplicate 30 mL cultures and each pellet was extracted with 5 mL pure methanol by shaking over night at 4 °C. Afterward, the extracts were centrifuged and the supernatant was evaporated via freeze drying. The dried extracts were then resuspended in 1 mL 50% methanol each, cleared via centrifugation and analyzed for production via LC-FT-MS (see Supplementary Tables S4a-S4d). Xanthomonin III variants were expressed accordingly as triplicate 500 mL M9 minimal medium cultures. Pellets were extracted with 25 mL MeOH by shaking over night at 4 °C. After centrifugation, the supernatant was evaporated via freeze drying and the dried extracts were resuspended in 1 mL 50% methanol and analyzed in a likewise manner to the Xanthomonin II variants (See Supplementary Table S5).

Mass Spectrometric Analysis. Mass spectrometric analysis of extracts was done using a LTQ-FT ultra instrument (Thermo Fisher Scientific) that was connected to a microbore 1100 HPLC system (Agilent). Separation of the extracts was achieved by using a CC 125/2 Nucleosil 300-8 C18 column (Macherey-Nagel) and a gradient of water/0.1% trifluoroacetic acid (solvent A) and MeCN/0.1% trifluoroacetic acid (solvent B) at 40 °C and a flow rate of 0.2 ml/min, while UV absorption was recorded at 215 nm for detection: Holding 2% B for 2 min, followed by a linear increase from 2% to 30% B in 18 min, a subsequent increase from 30% to 90% B in 15 min and holding 95% B for an additional 2 min. For each measurement, 100 µL of the respective samples were injected. Each sample was then analyzed for the presence of any possible truncation product, ranging from the full length product to the lone macrolactam ring. For quantification, the UV peak areas of all detectable truncations were integrated and added to determine the overall production. Relative production of variants was determined by comparison of their overall integral to the overall integral of the corresponding wild type samples. All data obtained this way is shown in Supplementary Tables S3-S5.

Analysis of thermal stability and protease assays of the purified lasso peptides was done with a 1100 series MSD (Hewlett-Packard) coupled with a microbore 1100 HPLC system (Agilent). For sample separation an EC 125/2 Nucleodur 100-3 C18ec column at room temperature and a flow rate of 0.2 mL/min was used. For xanthomonin I, the following gradient was used: Linear increase from 25% to 35% B in 15 min, followed by a linear gradient from 35% to 95% B in 2 min and holding 95% B for additional 5 min. For xanthomonin II, the following gradient was used: Linear increase from 20% to 35% B in 15 min, followed by a linear gradient from 35% to 95% B in 2 min and holding 95% B for additional 5 min. For xanthomonin III, the following gradient was used: Linear increase from 0% to 30% B in 15 min, followed by a linear gradient from 30% to 95% B in 2 min and holding 95% B for additional 5 min.

Tandem mass spectrometry was performed by employing collision-induced dissociation fragmentation within the linear ion trap using online LC-MS. As the singly charged ions were the dominant species for all xanthomonins, they were selected for fragmentation studies. In all instances, the energy for fragmentation was set to 35%.

Protease Assays of Xanthomonins I and II. For testing the proteolytic stability of Xanthomonins I and II, assays with chymotrypsin, proteinase K and carboxypeptidase Y were performed. These assays were performed over the course of 4 h at either 25 °C (chymostrypsin and carboxypeptidase Y) or 37 °C (proteinase K), using 10 μg of purified lasso peptide per assay. Chymotrypsin assays were done with 0.5 μg protease per reaction in a buffer containing 100 mM Tris-HCl and 10 mM $CaCl_2$ at pH 8.5. Proteinase K assays were done with 1 U proteinase K per reaction in a buffer containing 50 mM Tris-HCl at pH 7.5. Carboxypeptidase Y assays were done with 0.5 U carboxypeptidase Y per reaction in a buffer containing 50 mM MES and 1 mM $CaCl_2$ at pH 6.75. After theend of the incubation time, the samples were compared to an untreated sample via LC-MS.

Thermal Stability Assays of Xanthomonins I and II. For testing the thermal stability of xanthomonins I-II, 100 μL samples, containing each 10 μg of the corresponding lasso peptide in water, were incubated at 95 °C for different time intervals and were then compared to an untreated sample via LC-MS.

Crystallization and Data Collection. Saturated solutions of the -4 aa truncation of xanthomonin I and the -6 aa truncation of xanthomonin II in water were used for crystallization screenings using typical protein crystallization screens purchased from Qiagen. Crystals of xanthomonin I were obtained upon incubation at 277 K after mixing 300 nL peptide solution with 300 nL reservoir solution consisting of 8.5 mM cobalt chloride, 85 mM sodium acetate at pH 4.6, 0.85 M 1,6-hexanediol and 15% (V/V) glycerol. Crystals of xanthomonin II were obtained upon incubation at 291 K after mixing 300 nL peptide solution with 300 nL reservoir solution consisting of 2 M sodium chloride and 10% (w/V) polyethylene glycol 6000. Crystals were harvested from mother liquor and frozen without cryoprotectant in liquid nitrogen. Data was collected at 100 K at Bessy II (Helmholtz-Zentrum Berlin, Germany), Beamline 14.1 using a Dectris PILATUS detector.[4] In order to achieve high-resolution data for xanthomonin I, the wavelength was set to 0.800 Å and an oscillation range of 0.1° in order to improve the signal-to-noise ratio. Data processing and scaling was performed using the program XDS.[5] Data were prepared for solving the phase problem by direct methods using XPREP

Crystal Structure Elucidation of Xanthomonin I. Xanthomonin I crystallizes in the monoclinic space group $P2_1$ with cell constants a = 13.804(3) Å, b = 45.096(9) Å, c = 13.890(3) Å, β = 91.42°, V = 8644(3) $Å^3$, with two molecules, one bridging water molecule, 18 solvent water molecules and one molecule 1,6-hexanediol in the asymmetric unit, empirical formula $2(C_{65}H_{97}N_{17}O_{21}),C_6H_{12}O_2,19(H_2O)$, formula weight 3363.64 g/mol, calculated density 1.292 g/cm^3. The structure has been solved using dual

space direct methods[6] and refined using the full matrix least squares procedure,[7] non-hydrogen atoms of the peptide and the bridging water molecule were anisotropically refined. Hydrogen atoms have been mostly included at calculated positions; some N- or O-bonded hydrogen atoms involved in hydrogen bonds were located and isotropically refined. Solvent water oxygen atoms were refined with a common isotropic temperature factor without adding hydrogen atoms and with split positions.

The refinement of 2162 parameters converged to R1 = 0.0544 and R2 = 0.148 for all 24760 independent reflections. The largest peak and hole in the difference Fourier density map was 0.51 and -0.32 e $Å^{-3}$.

The coordinates of the crystal structure of the -4 aa truncation of xanthomonin I have been submitted to the Protein Data Bank (PDB) and assigned the accession number 4NAG.

NMR Structure Elucidation of Xanthomonin II. The sample for the NMR spectroscopic studies contained 2.8 mg of the -6 aa truncation of xanthomonin II in 250 µL DMSO-d_6 filled in a 3 mm Wilmad tube (Rototec Spintec). Spectra were recorded on a Bruker Avance 600 MHz spectrometer equipped with an inverse triple resonance ^{1}H-^{13}C-^{15}N probe with z-gradient. Temperature influence on the sample was studied by recording ^{1}H spectra at temperatures between 295 and 313 K with an increment of 5 K. The 2D experiments DQF-COSY,[8] TOCSY,[9] and NOESY[10] were performed at 305 K in the standard way for sequential assignments.[11] The TOCSY spectra were recorded with a mixing time of 120 ms, while NOESY spectra were recorded with mixing times of 100 and 300 ms. 1D spectra were acquired with 65536 data points, while 2D spectra were collected using 4096 points in the F_2 dimension and 512 increments in the F_1 dimension. Both 1D and 2D spectra were recorded with 32 transients and relaxation delays of 3.0 s. Chemical shifts were referenced to the rest signal of DMSO-d_6. All spectra were processed with Bruker TOPSPIN 3.1. NOE cross-peaks were analyzed within the program Sparky.[12]

Structure calculations were performed on the -6 aa truncation of xanthomonin II in DMSO-d_6 at 305 K. NOESY cross-peaks with a mixing time of 100 ms were used to create distance constraints. Constraints of the torsion angle φ were determined by analyzing the vicinal coupling constants $^{3}J_{HN\alpha}$, while those of χ^{1} were determined by fulfilling the stereospecific assignment on the basis of NOE contacts and $^{3}J_{\alpha\beta}$. These constraints were used in the simulated annealing protocol and the structure calculations were done with the program CYANA 2.1.[13] The coordinates of the 15 lowest energy structures of the -6 aa truncation of xanthomonin II have been submitted to the Protein Data Bank (PDB) and assigned the accession number 2MFV.

Antibacterial Assays. Spot-on-lawn assays were used to assess a potential antibacterial activity of xanthomonins I and II. These assays were performed with *Asticaccaulis excentricus*, *Bacillus subtilis*, *Burkholderia thailendensis*, *Burkholderia rhizoxinica*, *Caulobacter crescentus*, *Caulobacter* sp. K31, *Caulobacter segnis*, *Micrococcus flavus,* *Sphingobium japonicum*, *Sphingopyxis alaskensis* and *Xanthomonas gardneri*.

Soft agar was prepared by inoculation of soft agar medium (6 g/L agar) to an OD of ~0.01 with an culture of the corresponding bacterial strain during its exponential growth phase. These bacterial suspensions were then used for creating overlays on a 20 mL agar layer (15 g/L agar). Subsequently, several droplets of different volumes (0, 5, 10 or 20 µL of an 1 mg/mL stock solution of the respective lasso peptide) were placed on marked spots on the agar overlay. LB medium was used as both liquid and solid medium for all strains with exception of *Caulobacter crescentus* and *C.* sp. K31. For the latter strains, PYEGR medium was employed. The plates were analyzed for the presence of inhibition halos after both 1 and 2 days of incubation at 30 °C. In no case, the formation of inhibition halos was observed, meaning that all tested strains are not sensitive towards xanthomonins I and II in the tested concentration range.

Thermal Stability and Carboxypeptidase Y Assays of Xanthomonin II Variants and Xanthomonin III. For assessment of the topology and thermal stability of Xanthomonin II variants, 220 µL of the corresponding extracts were evaporated at 30 °C under reduced pressure and subsequently resuspended in 44 µL MES buffer (50 mM MES, 1 mM $CaCl_2$, pH 6.75). 22 µL of these solutions were incubated at 95 °C for 4 h. From both the untreated and thermally treated samples, 10 µL were mixed with 15 µL carboxypeptidase Y solution (100 U/mL, 50 mM MES, 1 mM $CaCl_2$, pH 6.75), incubated at 25 °C for 4 h and were afterward diluted with 75 µL water. As non-digested references, 10 µL of each untreated and thermally treated samples were diluted with 15 µL MES buffer and 75 µL water. As control, these experiments were also done with extracts of wild type xanthomonin II. All samples were then analyzed via LC-FT-MS.

Differentiation between a lasso peptide and an unthreaded branched cyclic peptide is possible via carboxypeptidase Y digestion. As the macrolactam ring shields the amino acids of the tail that are positioned inside of the ring against proteolytic digestion, a lasso peptide with a suitable plug amino acid at this position can only be degraded up or close to the plug amino acid. In contrast, a branched cyclic peptide can be digested up or close to the macrolactam ring. Analysis of the main products after carboxypeptidase Y digestion thus reveals the topology of the investigated variant. For a heat stable lasso peptide, carboxypeptidase Y digestion would show only degradation products up or close to the plug amino acid, independent of prior thermal treatment. On the other hand, a heat senstive lasso peptide

would show different degradation products, depending on the carboxypeptidase Y digestion being performed before or after thermal treatment.

In a similar manner, the thermal stability and topology of Xanthomonin III was assessed. In this case, 10 µL of a solution of pure -7 amino acid truncation of xanthomonin III in water (1 mg/mL) was used for each assay and as reference. If necessary, samples were incubated for 4 h at 95 °C. For carboxypeptidase Y assays, 25 µL of a carboxypeptidase Y solution (20 U/mL, 50 mM MES, 1 mM $CaCl_2$, pH 6.75) were added, the samples were then incubated at 25 °C for 4 h and afterward diluted with 75 µL water. Samples that should not be treated with carboxypeptidase Y were directly diluted with 25 µL MES buffer and 75 µL water. All samples were analyzed via LC-MS.

Supplementary Table S1a. Oligonucleotide primers for cloning of the investigated gene clusters. Introduced restriction sites are underlined.

name	sequence
XgaA1_NdeI_FP	GGATTAC CAT ATG AAC AGC AAC GAC ACC ACC CAC AGC G
XgaC_SpeI_RP	ATAT ACT AGT TCA GCC TTG ATG ACG GAC CCA GTT CTC GAC
XcitA_NdeI_FP	GG GAA TTC CAT ATG AAA ACG TTC GTC GAT AGC GCT CCG AGC CAG GAA GAC C
XcitC_HindIII_RP	ATAT AAG CTT CTA GCC GTG ATT GCG GAT CCA GTT CTC GAT CAT GCA CAG GTC

Supplementary Table S1b. Oligonucleotide primers for cluster optimization. Introduced restriction sites are underlined. Sequences introduced via mutagenesis are highlighted in bold.

name	sequence
pET41mcsKO_FP[a]	TAA TTG ATT AAT ACC TAG GCT GCT AAA CAA AGC CCG AAA GGA AGC TG
pET41mcsKO_RP[a]	CGG AGT GAT GGT GAT GGT GAT GAC CAG AAC CAC TAG TTC AGC
XgaBC-pET41a_ EcoRI_FP	ATATA GAA TTC ATG CAT CGC TAC ACA TTG CGC GAC GAC
XgaA2-pET41a_ KpnI_RP	ATAT GGT ACC TA GTC CTG GCT GAT GCC CAG GGT GGT G
LambdaTerm_Mcj Prom_KpnI_FP	ATAT GGT ACC AGC TTA ATT AGC TGA GCT TGG ACT CCT GTT GAT AG
LambdaTerm_Mcj Prom_EcoRI_RP	TATAT GAA TTC CAC TCC CTC CTG TAG TTA TGC CTT TAC ACT CAA TTG G
Xcit_LinkRBS_FP[a]	**AGAGGAGAAATTAACC** ATG GCT CGG CAT CTG AAC GAA GAC ATT TCG TAC TGC CAA C
Xcit_LinkRBS_RP[a]	TCA TTC TTC GGA GAT GCC GGC GAT GGG GGA CAT CCC

[a] Primers were 5'-phosphorylated

Supplementary Table S1c. Oligonucleotide primers for mutation of the xanthomonin II precursor peptide in between positions -2 to 9. All mutations were introduced using SLIM.[2-3] SLIM overhang regions are underlined. Mutated positions are highlighted in bold.

name	sequence
XgaA2_FP_-2-9	GGT GGC ATC ACC ACC CTG GGC ATC AGC CAG GAC TAG
XgaA2_RP_-2-9	**GCG** CAG GGC GGG CCA TTG GCA GGC GAA GAA ATG GGT GGC ATC ACC ACC CTG GGC ATC AGC CAG GAC TAG
XgaA2_T-2A_FPTail	**GCG** CAG GGC GGG CCA TTG GCA GGC GAA GAA ATG GGT GGC ATC ACC ACC CTG GGC ATC AGC CAG GAC TAG
XgaA2_T-2A_RPTail	CAT TTC TTC GCC TGC CAA TGG CCC GCC CTG **CGC** GTC CAG GCT TGC GAC CCC TAG CAC GAT GAG ATC CTG
XgaA2_T-2C_FPTail	**TGC** CAG GGC GGG CCA TTG GCA GGC GAA GAA ATG GGT GGC ATC ACC ACC CTG GGC ATC AGC CAG GAC TAG
XgaA2_T-2C_RPTail	CAT TTC TTC GCC TGC CAA TGG CCC GCC CTG **GCA** GTC CAG GCT TGC GAC CCC TAG CAC GAT GAG ATC CTG
XgaA2_T-2I_FPTail	**ATT** CAG GGC GGG CCA TTG GCA GGC GAA GAA ATG GGT GGC ATC ACC ACC CTG GGC ATC AGC CAG GAC TAG
XgaA2_T-2I_RPTail	CAT TTC TTC GCC TGC CAA TGG CCC GCC CTG **AAT** GTC CAG GCT TGC GAC CCC TAG CAC GAT GAG ATC CTG
XgaA2_T-2S_FPTail	**AGC** CAG GGC GGG CCA TTG GCA GGC GAA GAA ATG GGT GGC ATC ACC ACC CTG GGC ATC AGC CAG GAC TAG
XgaA2_T-2S_RPTail	CAT TTC TTC GCC TGC CAA TGG CCC GCC CTG **GCT** GTC CAG GCT TGC GAC CCC TAG CAC GAT GAG ATC CTG
XgaA2_T-2V_FPTail	**GTG** CAG GGC GGG CCA TTG GCA GGC GAA GAA ATG GGT GGC ATC ACC ACC CTG GGC ATC AGC CAG GAC TAG
XgaA2_T-2V_RPTail	CAT TTC TTC GCC TGC CAA TGG CCC GCC CTG **CAC** GTC CAG GCT TGC GAC CCC TAG CAC GAT GAG ATC CTG
XgaA2_P3A_FPTail	ACA CAG GGC GGG **GCG** TTG GCA GGC GAA GAA ATG GGT GGC ATC ACC ACC CTG GGC ATC AGC CAG GAC TAG
XgaA2_P3A_RPTail	CAT TTC TTC GCC TGC CAA **CGC** CCC GCC CTG TGT GTC CAG GCT TGC GAC CCC TAG CAC GAT GAG ATC CTG
XgaA2_E7A_FPTail	ACA CAG GGC GGG CCA TTG GCA GGC **GCG** GAA ATG GGT GGC ATC ACC ACC CTG GGC ATC AGC CAG GAC TAG
XgaA2_E7A_RPTail	CAT TTC **CGC** GCC TGC CAA TGG CCC GCC CTG TGT GTC CAG GCT TGC GAC CCC TAG CAC GAT GAG ATC CTG
XgaA2_E7D_E8A_FPTail	ACA CAG GGC GGG CCA TTG GCA GGC **GAT GCG** ATG GGT GGC ATC ACC ACC CTG GGC ATC AGC CAG GAC TAG
XgaA2_E7D_E8A_RPTail	CAT **CGC ATC** GCC TGC CAA TGG CCC GCC CTG TGT GTC CAG GCT TGC GAC CCC TAG CAC GAT GAG ATC CTG
XgaA2_E8A_FPTail	ACA CAG GGC GGG CCA TTG GCA GGC GAA **GCG** ATG GGT GGC ATC ACC ACC CTG GGC ATC AGC CAG GAC TAG
XgaA2_E8A_RPTail	CAT **CGC** TTC GCC TGC CAA TGG CCC GCC CTG TGT GTC CAG GCT TGC GAC CCC TAG CAC GAT GAG ATC CTG
XgaA2_E7A_E8D_FPTail	ACA CAG GGC GGG CCA TTG GCA GGC **GCG GAT** ATG GGT GGC ATC ACC ACC CTG GGC ATC AGC CAG GAC TAG
XgaA2_E7A_E8D_RPTail	CAT **ATC CGC** GCC TGC CAA TGG CCC GCC CTG TGT GTC CAG GCT TGC GAC CCC TAG CAC GAT GAG ATC CTG

Supplementary Table S1d. Oligonucleotide primers for mutation of the xanthomonin II precursor peptide in between positions 10 to 20. All mutations were introduced using SLIM.[2-3] SLIM overhang regions are underlined. Mutated positions are highlighted in bold.

name	sequence
XgaA2_FP_10-20	TAG GTA CCA GCT TAA TTA GCT GAG CTT GGA CTC CTG TTG ATA G
XgaA2_RP_10-20	CAT TTC TTC GCC TGC CAA TGG CCC GCC CTG TGT GTC CAG
XgaA2_G10A_FPTail	**GCG** GGC ATC ACC ACC CTG GGC ATC AGC CAG GAC TAG GTA CCA GCT TAA TTA GCT GAG CTT GGA CTC CTG TTG ATA G
XgaA2_G10A_RPTail	GTC CTG GCT GAT GCC CAG GGT GGT GAT GCC **CGC** CAT TTC TTC GCC TGC CAA TGG CCC GCC CTG TGT GTC CAG
XgaA2_G10L_FPTail	**CTG** GGC ATC ACC ACC CTG GGC ATC AGC CAG GAC TAG GTA CCA GCT TAA TTA GCT GAG CTT GGA CTC CTG TTG ATA G
XgaA2_G10L_RPTail	GTC CTG GCT GAT GCC CAG GGT GGT GAT GCC **CAG** CAT TTC TTC GCC TGC CAA TGG CCC GCC CTG TGT GTC CAG
XgaA2_G10F_FPTail	**TTT** GGC ATC ACC ACC CTG GGC ATC AGC CAG GAC TAG GTA CCA GCT TAA TTA GCT GAG CTT GGA CTC CTG TTG ATA G
XgaA2_G10F_RPTail	GTC CTG GCT GAT GCC CAG GGT GGT GAT GCC **AAA** CAT TTC TTC GCC TGC CAA TGG CCC GCC CTG TGT GTC CAG
XgaA2_G11A_FPTail	GGT **GCG** ATC ACC ACC CTG GGC ATC AGC CAG GAC TAG GTA CCA GCT TAA TTA GCT GAG CTT GGA CTC CTG TTG ATA G
XgaA2_G11A_RPTail	GTC CTG GCT GAT GCC CAG GGT GGT GAT **CGC** ACC CAT TTC TTC GCC TGC CAA TGG CCC GCC CTG TGT GTC CAG
XgaA2_G11L_FPTail	GGT **CTG** ATC ACC ACC CTG GGC ATC AGC CAG GAC TAG GTA CCA GCT TAA TTA GCT GAG CTT GGA CTC CTG TTG ATA G
XgaA2_G11L_RPTail	GTC CTG GCT GAT GCC CAG GGT GGT GAT **CAG** ACC CAT TTC TTC GCC TGC CAA TGG CCC GCC CTG TGT GTC CAG
XgaA2_G11F_FPTail	GGT **TTT** ATC ACC ACC CTG GGC ATC AGC CAG GAC TAG GTA CCA GCT TAA TTA GCT GAG CTT GGA CTC CTG TTG ATA G
XgaA2_G11F_RPTail	GTC CTG GCT GAT GCC CAG GGT GGT GAT **AAA** ACC CAT TTC TTC GCC TGC CAA TGG CCC GCC CTG TGT GTC CAG
XgaA2_I12A_FPTail	GGT GGC **GCG** ACC ACC CTG GGC ATC AGC CAG GAC TAG GTA CCA GCT TAA TTA GCT GAG CTT GGA CTC CTG TTG ATA G
XgaA2_I12A_RPTail	GTC CTG GCT GAT GCC CAG GGT GGT **CGC** GCC ACC CAT TTC TTC GCC TGC CAA TGG CCC GCC CTG TGT GTC CAG
XgaA2_I12A_T13A_FPTail	GGT GGC **GCG GCC** ACC CTG GGC ATC AGC CAG GAC TAG GTA CCA GCT TAA TTA GCT GAG CTT GGA CTC CTG TTG ATA G
XgaA2_I12A_T13A_RPTail	GTC CTG GCT GAT GCC CAG GGT **GGC CGC** GCC ACC CAT TTC TTC GCC TGC CAA TGG CCC GCC CTG TGT GTC CAG
XgaA2_I12A_T13A_T14A_FPTail	GGT GGC **GCG GCC GCA** CTG GGC ATC AGC CAG GAC TAG GTA CCA GCT TAA TTA GCT GAG CTT GGA CTC CTG TTG ATA G
XgaA2_I12A_T13A_T14A_RPTail	GTC CTG GCT GAT GCC CAG **TGC GGC CGC** GCC ACC CAT TTC TTC GCC TGC CAA TGG CCC GCC CTG TGT GTC CAG
XgaA2_I12A_T13A_T14A_L15A_FPTail	GGT GGC **GCG GCC GCA GCG** GGC ATC AGC CAG GAC TAG GTA CCA GCT TAA TTA GCT GAG CTT GGA CTC CTG TTG ATA G
XgaA2_I12A_T13A_T14A_L15A_RPTail	GTC CTG GCT GAT GCC **CGC TGC GGC CGC** GCC ACC CAT TTC TTC GCC TGC CAA TGG CCC GCC CTG TGT GTC CAG

Supplementary Table S1e. Oligonucleotide primers for mutation of the xanthomonin II precursor peptide in between positions 10 to 20. All mutations were introduced using SLIM.[2-3] SLIM overhang regions are underlined. Mutated positions are highlighted in bold.

name	sequence
XgaA2_T13A_T14A_L15A_FPTail	GGT GGC ATC **GCC GCA GCG** GGC ATC AGC CAG GAC TAG GTA CCA GCT TAA TTA GCT GAG CTT GGA CTC CTG TTG ATA G
XgaA2_T13A_T14A_L15A_RPTail	GTC CTG GCT GAT GCC **CGC TGC GGC** GAT GCC ACC CAT TTC TTC GCC TGC CAA TGG CCC GCC CTG TGT GTC CAG
XgaA2_I12L_T13A_T14A_L15A_FPTail	GGT GGC **CTG GCC GCA GCG** GGC ATC AGC CAG GAC TAG GTA CCA GCT TAA TTA GCT GAG CTT GGA CTC CTG TTG ATA G
XgaA2_I12L_T13A_T14A_L15A_RPTail	GTC CTG GCT GAT GCC **CGC TGC GGC CAG** GCC ACC CAT TTC TTC GCC TGC CAA TGG CCC GCC CTG TGT GTC CAG
XgaA2_I12V_T13A_T14A_L15A_FPTail	GGT GGC **GTG GCC GCA GCG** GGC ATC AGC CAG GAC TAG GTA CCA GCT TAA TTA GCT GAG CTT GGA CTC CTG TTG ATA G
XgaA2_I12V_T13A_T14A_L15A_RPTail	GTC CTG GCT GAT GCC **CGC TGC GGC CAC** GCC ACC CAT TTC TTC GCC TGC CAA TGG CCC GCC CTG TGT GTC CAG
XgaA2_I12S_T13A_T14A_L15A_FPTail	GGT GGC **AGC GCC GCA GCG** GGC ATC AGC CAG GAC TAG GTA CCA GCT TAA TTA GCT GAG CTT GGA CTC CTG TTG ATA G
XgaA2_I12S_T13A_T14A_L15A_RPTail	GTC CTG GCT GAT GCC **CGC TGC GGC GCT** GCC ACC CAT TTC TTC GCC TGC CAA TGG CCC GCC CTG TGT GTC CAG
XgaA2_I12T_T13A_T14A_L15A_FPTail	GGT GGC **ACC GCC GCA GCG** GGC ATC AGC CAG GAC TAG GTA CCA GCT TAA TTA GCT GAG CTT GGA CTC CTG TTG ATA G
XgaA2_I12T_T13A_T14A_L15A_RPTail	GTC CTG GCT GAT GCC **CGC TGC GGC GGT** GCC ACC CAT TTC TTC GCC TGC CAA TGG CCC GCC CTG TGT GTC CAG

Supplementary Table S1f. Oligonucleotide primers for mutation of the xanthomonin III precursor peptide in between positions 10 to 17. All mutations were introduced using SLIM.[2-3] SLIM overhang regions are underlined. Mutated positions are highlighted in bold.

name	sequence
Xcit_10-17_SLIM_FP	TCC GAA GAA TGA AGA GGA GAA ATT AAC CAT GGC TCG GCA TC
Xcit_10-17_SLIM_RP	ATT CAC CTC ACC GGC TCC CGC GCC GCC CTG GG
Xcit_I14A_FPTail	GGG ATG TCC CCC **GCG** GCC GGC ATC TCC GAA GAA TGA AGA GGA GAA ATT AAC CAT GGC TCG GCA TC
Xcit_I14A_RPTail	GAT GCC GGC **CGC** GGG GGA CAT CCC ATT CAC CTC ACC GGC TCC CGC GCC GCC CTG GG
Xcit_P13A_I14A_FPTail	GGG ATG TCC **GCA GCG** GCC GGC ATC TCC GAA GAA TGA AGA GGA GAA ATT AAC CAT GGC TCG GCA TC
Xcit_P13A_I14A_RPTail	GAT GCC GGC **CGC TGC** GGA CAT CCC ATT CAC CTC ACC GGC TCC CGC GCC GCC CTG GG
Xcit_S12A_P13A_I14A_FPTail	GGG ATG **GCT GCA GCG** GCC GGC ATC TCC GAA GAA TGA AGA GGA GAA ATT AAC CAT GGC TCG GCA TC
Xcit_S12A_P13A_I14A_RPTail	GAT GCC GGC **CGC TGC AGC** CAT CCC ATT CAC CTC ACC GGC TCC CGC GCC GCC CTG GG
Xcit_M11A_S12A_P13A_I14A_FPTail	GGG **GCC GCT GCA GCG** GCC GGC ATC TCC GAA GAA TGA AGA GGA GAA ATT AAC CAT GGC TCG GCA TC
Xcit_M11A_S12A_P13A_I14A_RPTail	GAT GCC GGC **CGC TGC AGC GGC** CCC ATT CAC CTC ACC GGC TCC CGC GCC GCC CTG GG

Supplementary Table S2. M9 Vitamin Mix.

component	amount
choline chloride	1.0 g
folic acid	1.0 g
pantothenic acid	1.0 g
nicotinamide	1.0 g
myo-inositol	2.0 g
pyridoxal hydrochloride	1.0 g
thiamine	1.0 g
riboflavin	0.1 g
disodium adenosine 5'-triphosphate	0.3 g
biotin	0.2 g
	ad 300 mL ddH_2O[a]

[a]10 M NaOH was slowly added to the resulting suspension until everything was dissolved (pH ~12). Afterward the solution was sterile filtered and stored at 4 °C until needed.

Supplementary Table S3a. Production of different constructs generated for production optimization of xanthomonins I and II. Truncations that were detectable by MS but not UV are marked with a "-".

name	sample	UV peak areas of all detectable compounds	sum of UV peak areas	relative production
xgaA1A2BC pET41a	1	8.8E+6 (xanthomonin I, -4 aa) - (xanthomonin II, -6 aa)	8.8E+6	
	2	9.1E+6 (xanthomonin I, -4 aa) - (xanthomonin II, -6 aa)	9.1E+6	
	3	8.6E+6 (xanthomonin I, -4 aa) - (xanthomonin II, -6 aa)	8.6E+6	
	mean	**8.8E+6 (xanthomonin I, -4 aa)** **- (xanthomonin II, -6 aa)**	**8.8E+6**	**1.00 ± 0.02**
*xgaA1*_T-P_*BC* pET41a	1	1.4E+6 (xanthomonin I, full) 2.0E+6 (xanthomonin I, -3 aa) 3.2E+8 (xanthomonin I, -4 aa)	3.2E+8	
	2	1.4E+6 (xanthomonin I, full) 2.1E+6 (xanthomonin I, -3 aa) 3.2E+8 (xanthomonin I, -4 aa)	3.3E+8	
	3	1.4E+6 (xanthomonin I, full) 2.2E+6 (xanthomonin I, -3 aa) 3.4E+8 (xanthomonin I, -4 aa)	3.4E+8	
	mean	**1.4E+6 (xanthomonin I, full)** **2.1E+6 (xanthomonin I, -3 aa)** **3.3E+8 (xanthomonin I, -4 aa)**	**3.3E+8**	**37.27 ± 0.75**
*xgaA2*_T-P_*BC* pET41a	1	1.6E+7 (xanthomonin II, -4 aa) 6.0E+6 (xanthomonin II, -5 aa) 6.6E+7 (xanthomonin II, -6 aa)	8.8E+7	
	2	2.1E+7 (xanthomonin II, -4 aa) 8.5E+6 (xanthomonin II, -5 aa) 8.6E+7 (xanthomonin II, -6 aa)	1.2E+8	
	3	1.5E+7 (xanthomonin II, -4 aa) 5.3E+6 (xanthomonin II, -5 aa) 6.0E+7 (xanthomonin II, -6 aa)	8.0E+7	
	mean	**1.7E+7 (xanthomonin II, -4 aa)** **6.6E+6 (xanthomonin II, -5 aa)** **7.1E+7 (xanthomonin II, -6 aa)**	**9.5E+7**	**10.70 ± 1.69**

Supplementary Table S3b. Production of different constructs generated for production optimization of xanthomonins III. Truncations that were detectable by MS but not UV are marked with a "-".

name	sample	UV peak areas of all detectable compounds	sum of UV peak areas	relative production
xcitABC pET41a[a]	1	1.4E+8 (xanthomonin III, -4 aa/-5 aa) - (xanthomonin III, -6 aa) 1.0E+8 (xanthomonin III, -7 aa)	2.4E+8	
	2	9.4E+7 (xanthomonin III, -4 aa/-5 aa) - (xanthomonin III, -6 aa) 8.8E+7 (xanthomonin III, -7 aa)	1.8E+8	
	3	1.1E+8 (xanthomonin III, -4 aa/-5 aa) - (xanthomonin III, -6 aa) 8.0E+7 (xanthomonin III, -7 aa)	1.8E+8	
	mean	**1.1E+8 (xanthomonin III, -4 aa/-5 aa)** **- (xanthomonin III, -6 aa)** **8.9E+7 (xanthomonin III, -7 aa)**	**2.0E+8**	**1.00 ± 0.13**
xcitA_RBS_BC pET41a[a]	1	- (xanthomonin III, full) 4.2E+8 (xanthomonin III, -4 aa/-5 aa) - (xanthomonin III, -6 aa) 2.6E+8 (xanthomonin III, -7 aa)	6.8E+8	
	2	- (xanthomonin III, full) 4.9E+8 (xanthomonin III, -4 aa/-5 aa) - (xanthomonin III, -6 aa) 3.3E+8 (xanthomonin III, -7 aa)	8.2E+8	
	3	- (xanthomonin III, full) 4.3E+8 (xanthomonin III, -4 aa/-5 aa) - (xanthomonin III, -6 aa) 2.4E+8 (xanthomonin III, -7 aa)	6.7E+8	
	mean	**- (xanthomonin III, full)** **4.5E+8 (xanthomonin III, -4 aa/-5 aa)** **- (xanthomonin III, -6 aa)** **2.8E+8 (xanthomonin III, -7 aa)**	**7.2E+8**	**3.62 ± 0.34**

[a] UV signals of -4 aa and -5 aa truncations strongly overlap. The intensity of their MS signals suggests an approximately equimolar ratio.

Supplementary Table S4a. Production of different xanthomonin II variants. Truncations that were detectable by MS but not UV are marked with a "-".

name	sample	UV peak areas of all detectable compounds	sum of UV peak areas	relative production
wild type	1	1.6E+7 (-4 aa), 6.0E+6 (-5 aa), 6.6E+7 (-6 aa)	8.8E+7	
	2	2.1E+7 (-4 aa), 8.5E+6 (-5 aa), 8.6E+7 (-6 aa)	1.2E+8	
	3	1.5E+7 (-4 aa), 5.3E+6 (-5 aa), 6.0E+7 (-6 aa)	8.0E+7	
	mean	**1.7E+7 (-4 aa), 6.6E+6 (-5 aa), 7.1E+7 (-6 aa)**	**9.5E+7**	**1.000 ± 0.158**
T-2A	1	- (-4 aa), - (-5 aa), 3.6E+5 (-6 aa)	3.6E+5	
	2	- (-4 aa), - (-5 aa), 4.8E+5 (-6 aa)	4.8E+5	
	3	- (-4 aa), - (-5 aa), 4.4E+5 (-6 aa)	4.4E+5	
	mean	**- (-4 aa), - (-5 aa), 4.3E+5 (-6 aa)**	**4.3E+5**	**0.005 ± 0.001**
T-2C	1	2.6E+6 (-4 aa), 1.4E+6 (-5 aa), 9.3E+6 (-6 aa)	1.3E+7	
	2	2.2E+6 (-4 aa), 8.7E+5 (-5 aa), 8.6E+6 (-6 aa)	1.2E+7	
	3	2.1E+6 (-4 aa), 9.6E+5 (-5 aa), 8.2E+6 (-6 aa)	1.1E+7	
	mean	**2.3E+6 (-4 aa), 1.1E+6 (-5 aa), 8.7E+6 (-6 aa)**	**1.2E+7**	**0.128 ± 0.009**
T-2I	1	4.9E+5 (-4 aa), - (-5 aa), 5.3E+5 (-6 aa)	1.0E+6	
	2	4.5E+5 (-4 aa), - (-5 aa), 5.3E+5 (-6 aa)	9.9E+5	
	3	4.3E+5 (-4 aa), - (-5 aa), 5.1E+5 (-6 aa)	9.5E+5	
	mean	**4.6E+5 (-4 aa), - (-5 aa), 5.2E+5 (-6 aa)**	**9.8E+5**	**0.010 ± 0.001**
T-2S		4.9E+5 (-4 aa), - (-5 aa), 4.3E+5 (-6 aa)	9.2E+5	
		4.9E+5 (-4 aa), - (-5 aa), 4.6E+5 (-6 aa)	9.6E+5	
		5.0E+5 (-4 aa), - (-5 aa), 4.1E+5 (-6 aa)	9.1E+5	
		4.9E+5 (-4 aa), - (-5 aa), 4.3E+5 (-6 aa)	**9.3E+5**	**0.010 ± 0.001**
T-2V		7.3E+5 (-4 aa), - (-5 aa), 1.1E+6 (-6 aa)	1.8E+6	
		4.9E+5 (-4 aa), - (-5 aa), 9.8E+5 (-6 aa)	1.5E+6	
		3.6E+5 (-4 aa), - (-5 aa), 9.2E+5 (-6 aa)	1.3E+6	
		4.9E+5 (-4 aa), - (-5 aa), 1.0E+6 (-6 aa)	**1.5E+6**	**0.016 ± 0.021**
P3A	1	- (full), 4.7E+6 (-4 aa), 2.0E+6 (-5 aa), 4.2E+7 (-6 aa)	4.8E+7	
	2	- (full), 3.6E+6 (-4 aa), 9.9E+5 (-5 aa), 4.0E+7 (-6 aa)	4.5E+7	
	3	- (full), 4.8E+6 (-4 aa), 1.7E+6 (-5 aa), 4.1E+7 (-6 aa)	4.8E+7	
	mean	**- (full), 4.7E+6 (-4 aa), 1.6E+6 (-5 aa), 4.1E+7 (-6 aa)**	**4.7E+7**	**0.495 ± 0.014**
E7A	1	-	-	
	2	-	-	
	3	-	-	
	mean	**-**	**-**	**no production**
E7D E8A	1	-	-	
	2	-	-	
	3	-	-	
	mean	**-**	**-**	**no production**

Supplementary Table S4b. Production of different xanthomonin II variants. Truncations that were detectable by MS but not UV are marked with a "-".

name	sample	UV peak areas of all detectable compounds	sum of UV peak areas	relative production
E8A	1	1.3E+7 (-4 aa), 6.7E+6 (-5 aa), 1.2E+7 (-6 aa)	3.1E+7	
	2	2.0E+7 (-4 aa), 1.0E+7 (-5 aa), 1.2E+7 (-6 aa)	4.2E+7	
	3	1.9E+7 (-4 aa), 9.2E+6 (-5 aa), 1.1E+7 (-6 aa)	3.9E+7	
	mean	**1.7E+7 (-4 aa), 8.6E+6 (-5 aa), 1.2E+7 (-6 aa)**	**3.7E+7**	**0.395 ± 0.048**
E7A E8D	1	-	-	
	2	-	-	
	3	-	-	
	mean	**-**	**-**	**no production**
G10A	1	- (-12 aa), - (-13 aa)	-	
	2	- (-12 aa), - (-13 aa)	-	
	3	- (-12 aa), - (-13 aa)	-	
	mean	**- (-12 aa), - (-13 aa)**	**-**	**no UV**
G10L	1	- (-10 aa), - (-11 aa), - (-12 aa), - (-13 aa)	-	
	2	- (-10 aa), - (-11 aa), - (-12 aa), - (-13 aa)	-	
	3	- (-10 aa), - (-11 aa), - (-12 aa), - (-13 aa)	-	
	mean	**- (-10 aa), - (-11 aa), - (-12 aa), - (-13 aa)**	**-**	**no UV**
G10F	1	- (-10 aa), - (-11 aa), - (-12 aa), - (-13 aa)	-	
	2	- (-10 aa), - (-11 aa), - (-12 aa), - (-13 aa)	-	
	3	- (-10 aa), - (-11 aa), - (-12 aa), - (-13 aa)	-	
	mean	**- (-10 aa), - (-11 aa), - (-12 aa), - (-13 aa)**	**-**	**no UV**
G11A	1	6.6E+4 (-4 aa), - (-5 aa), 1.1E+6 (-6 aa), - (-10 aa), - (-13 aa)	1.2 E6	
	2	5.7E+4 (-4 aa), - (-5 aa), 9.7E+5 (-6 aa), - (-10 aa), - (-13 aa)	1.0 E6	
	3	5.9E+4 (-4 aa), - (-5 aa), 9.2E+5 (-6 aa), - (-10 aa), - (-13 aa)	9.8 E5	
	mean	6.0E+4 (-4 aa), - (-5 aa), 1.0E+6 (-6 aa), - (-10 aa), - (-13 aa)	1.1 E6	**0.011 ± 0.001**
G11L	1	- (-5 aa), - (-6 aa), - (-9 aa), - (-10 aa), - (-11 aa), - (-12 aa), - (-13 aa)	-	
	2	- (-5 aa), - (-6 aa), - (-9 aa), - (-10 aa), - (-11 aa), - (-12 aa), - (-13 aa)	-	
	3	- (-5 aa), - (-6 aa), - (-9 aa), - (-10 aa), - (-11 aa), - (-12 aa), - (-13 aa)	-	
	mean	**- (-5 aa), - (-6 aa), - (-9 aa), - (-10 aa), - (-11 aa), - (-12 aa), - (-13 aa)**	**-**	**no UV**
G11F	1	- (-6 aa), - (10 aa), - (-12 aa), - (-13 aa)	-	
	2	- (-6 aa), - (10 aa), - (-12 aa), - (-13 aa)	-	
	3	- (-6 aa), - (10 aa), - (-12 aa), - (-13 aa)	-	
	mean	**- (-6 aa), - (10 aa), - (-12 aa), - (-13 aa)**	**-**	**no UV**

Supplementary Table S4c. Production of different xanthomonin II variants. Truncations that were detectable by MS but not UV are marked with a "-".

name	sample	UV peak areas of all detectable compounds	sum of UV peak areas	relative production
I12A	1	6.5E+6 (-4 aa), 2.8E+7 (-5 aa), 1.0E+8 (-6 aa)	1.4E+8	
	2	6.4E+6 (-4 aa), 2.7E+7 (-5 aa), 9.6E+7 (-6 aa)	1.3E+8	
	3	6.8E+6 (-4 aa), 2.7E+7 (-5 aa), 9.3E+7 (-6 aa)	1.3E+8	
	mean	**6.6E+6 (-4 aa), 2.7E+7 (-5 aa), 9.7E+7 (-6 aa)**	**1.3E+8**	**1.382 ± 0.040**
I12A T13A	1	2.5E+7 (-4 aa), 4.5E+7 (-5 aa), 2.6E+7 (-6 aa), - (-11 aa), - (-12 aa), - (-13 aa)	9.6E+7	
	2	2.5E+7 (-4 aa), 4.4E+7 (-5 aa), 2.6E+7 (-6 aa), - (-11 aa), - (-12 aa), - (-13 aa)	9.4E+7	
	3	2.5E+7 (-4 aa), 4.5E+7 (-5 aa), 2.6E+7 (-6 aa), - (-11 aa), - (-12 aa), - (-13 aa)	9.6E+7	
	mean	**2.5E+7 (-4 aa), 4.5E+7 (-5 aa), 2.6E+7 (-6 aa), - (-11 aa), - (-12 aa), - (-13 aa)**	**9.5E+7**	**1.008 ± 0.010**
12A T13A T14A	1	3.0E+7 (-4 aa), 5.3E+7 (-5 aa), - (-6 aa), - (-8 aa), - (-10 aa), - (-11 aa), - (-12 aa), - (-13 aa)	3.6E+7	
	2	3.0E+7 (-4 aa), 5.4E+7 (-5 aa), - (-6 aa), - (-8 aa), - (-10 aa), - (-11 aa), - (-12 aa), - (-13 aa)	3.6E+7	
	3	3.0E+7 (-4 aa), 6.7E+7 (-5 aa), - (-6 aa), - (-8 aa), - (-10 aa), - (-11 aa), - (-12 aa), - (-13 aa)	3.7E+7	
	mean	**3.0E+7 (-4 aa), 5.8E+7 (-5 aa), - (-6 aa), - (-8 aa), - (-10 aa), - (-11 aa), - (-12 aa), - (-13 aa)**	**3.6E+7**	**0.380 ± 0.005**
I12A T13A T14A L15A	1	- (full), - (-4 aa), - (-8 aa), 5.0E+6 (-10 aa), - (-11 aa), 6.9E+5 (-12 aa), - (13 aa)	5.7E+6	
	2	- (full), - (-4 aa), - (-8 aa), 4.1E+6 (-10 aa), - (-11 aa), 5.2E+5 (-12 aa), - (13 aa)	4.7E+6	
	3	- (full), - (-4 aa), - (-8 aa), 4.0E+6 (-10 aa), - (-11 aa), 5.7E+5 (-12 aa), - (13 aa)	4.6E+6	
	mean	**- (full), - (-4 aa), - (-8 aa), 4.4E+6 (-10 aa), - (-11 aa), 5.9E+5 (-12 aa), - (13 aa)**	**5.0E+6**	**0.053 ± 0.005**
T13A T14A L15A	1	9.7E+7 (-4 aa), 3.6E+6 (-5 aa), 1.7E+7 (-6 aa)	1.2E+8	
	2	1.0E+8 (-4 aa), 3.9E+6 (-5 aa), 1.8E+7 (-6 aa)	1.2E+8	
	3	1.0E+8 (-4 aa), 3.8E+6 (-5 aa), 1.8E+7 (-6 aa)	1.2E+8	
	mean	**1.0E+8(-4 aa), 3.8E+6 (-5 aa), 1.8E+7 (-6 aa)**	**1.2E+8**	**1.280 ± 0.028**
I12L T13A T14A L15A	1	- (full), - (-3 aa), 1.0E+8 (-4 aa), 8.3E+6 (-5 aa), 5.4E+6 (-6 aa)	1.1E+8	
	2	- (full), - (-3 aa), 1.1E+8 (-4 aa), 7.7E+6 (-5 aa), 6.6E+6 (-6 aa)	1.2E+8	
	3	- (full), - (-3 aa), 1.1E+8 (-4 aa), 7.3E+6 (-5 aa), 6.1E+6 (-6 aa)	1.2E+8	
	mean	**- (full), - (-3 aa), 1.1E+8 (-4 aa), 7.8E+6 (-5 aa), 6.0E+6 (-6 aa)**	**1.2E+8**	**1.253 ± 0.032**
I12V T13A T14A L15A	1	- (full), 6.9E+7 (-4 aa), 2.0E+6 (-5 aa), 1.6E+7 (-6 aa)	8.7E+7	
	2	- (full), 7.4E+7 (-4 aa), 2.8E+6 (-5 aa), 1.7E+7 (-6 aa)	9.4E+7	
	3	- (full), 7.2E+7 (-4 aa), 3.7E+6 (-5 aa), 1.7E+7 (-6 aa)	9.3E+7	
	mean	**- (full), 7.1E+7 (-4 aa), 2.8E+6 (-5 aa), 176E+7 (-6 aa)**	**9.1E+7**	**0.964 ± 0.032**

Supplementary Table S4d. Production of different xanthomonin II variants. Truncations that were detectable by MS but not UV are marked with a "-".

name	sample	UV peak areas of all detectable compounds	sum of UV peak areas	relative production
I12S T13A T14A L15A[a]	1	- (full), 1.2E+7 (-4 aa/-10 aa), - (-8 aa), - (-11 aa), - (12 aa), - (-13 aa)	1.2E+7	
	2	- (full), 9.3E+6 (-4 aa/-10 aa), - (-8 aa), - (-11 aa), - (12 aa), - (-13 aa)	9.3E+6	
	3	- (full), 1.2E+7 (-4 aa/-10 aa), - (-8 aa), - (-11 aa), - (12 aa), - (-13 aa)	1.2E+7	
	mean	**- (full), 1.1E+7 (-4 aa/-10 aa), - (-8 aa), - (-11 aa), - (12 aa), - (-13 aa)**	**1.1E+7**	**0.119 ± 0.014**
I12T T13A T14A L15A	1	1.1E+8 (-4 aa), 1.8E+6 (-5 aa), 8.4E+6 (-6 aa)	1.2E+8	
	2	1.0E+8 (-4 aa), 3.8E+6 (-5 aa), 7.9E+6 (-6 aa)	1.1E+8	
	3	1.0E+8 (-4 aa), 2.4E+6 (-5 aa), 7.0E+6 (-6 aa)	1.1E+8	
	mean	**1.0E+8 (-4 aa), 2.7E+6 (-5 aa), 7.7E+6 (-6 aa)**	**1.2E+8**	**1.212 ± 0.037**

[a] UV signals of -4 aa and -10 aa truncations strongly overlap. The intensity of their MS signals suggests an approximately equimolar ratio.

Supplementary Table S5. Production of different xanthomonin III variants. Truncations that were detectable by MS but not UV are marked with a "-".

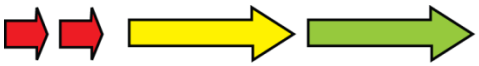
name	sample	UV peak areas of all detectable compounds	sum of UV peak areas	relative production
wild type[a]	1	- (full), 4.2E+8 (-4 aa/-5 aa), - (-6 aa), 2.6E+8 (-7 aa)	6.8E+8	
	2	- (full), 4.9E+8 (-4 aa/-5 aa), - (-6 aa), 3.3E+8 (-7 aa)	8.2E+8	
	3	- (full), 4.3E+8 (-4 aa/-5 aa), - (-6 aa), 2.4E+8 (-7 aa)	6.7E+8	
	mean	**- (full), 4.5E+8 (-4 aa/-5 aa), - (-6 aa), 2.8E+8 (-7 aa)**	**7.2E+8**	**1.000 ± 0.093**
I14A[b]	1	- (full), - (-4 aa) - (-5 aa), 8.0E+8 (-6 aa/-7 aa)	8.0E+8	
	2	- (full), - (-4 aa) - (-5 aa), 7.4E+8 (-6 aa/-7 aa)	7.4E+8	
	3	- (full), - (-4 aa) - (-5 aa), 7.1E+8 (-6 aa/-7 aa)	7.1E+8	
	mean	**- (full), - (-4 aa) - (-5 aa), 7.5E+8 (-6 aa/-7 aa)**	**7.5E+8**	**1.035 ± 0.055**
P13A I14A[b]	1	- (-4 aa) - (-5 aa), 5.8E+8 (-6 aa/-7 aa)	5.8E+8	
	2	- (-4 aa) - (-5 aa), 6.1E+8 (-6 aa/-7 aa)	6.1E+8	
	3	- (-4 aa) - (-5 aa), 6.8E+8 (-6 aa/-7 aa)	6.8E+8	
	mean	**- (-4 aa) - (-5 aa), 6.2E+8 (-6 aa/-7 aa)**	**6.2E+8**	**0.861 ± 0.054**
S12A P13A I14A[b]	1	- (full), - (-4 aa) - (-5 aa), 8.5E+8 (-6 aa/-7 aa)	8.5E+8	
	2	- (full), - (-4 aa) - (-5 aa), 8.3E+8 (-6 aa/-7 aa)	8.3E+8	
	3	- (full), - (-4 aa) - (-5 aa), 8.7E+8 (-6 aa/-7 aa)	8.7E+8	
	mean	**- (full), - (-4 aa) - (-5 aa), 8.5E+8 (-6 aa/-7 aa)**	**8.5E+8**	**1.173 ± 0.027**
M11A S12A P13A I14A	1	-	-	
	2	-	-	
	3	-	-	
	mean	**-**	**-**	**no production**

[a] UV signals of -4 aa and -5 aa truncations strongly overlap. The intensity of their MS signals suggests an approximately equimolar ratio.

[b] UV signals of -6 aa and -7 aa truncations strongly overlap. The intensity of their MS signals suggests an approximately equimolar ratio.

Structural Investigations and Analysis of the Xanthomonins

Comparison of All Previously Identified Lasso Peptide Precursors Containing Seven-Residue Macrolactam Rings. In a previously reported B-protein centric genome mining study,[14] we were able to identify a total of seven different putative lasso peptide precursor peptides that carry a glutamate residue suitable for macrolactam ring formation at position 7 of the proposed lasso peptide sequences. These precursor peptides were found in *Novopshingobium* sp. PP1Y, *Rhodanobacter* sp. 115, *Rhodanobacter* sp. 116-2, *Rhodanobacter* sp. 2APBS1, *Xanthomonas citri pv. mangiferaindicae* and *Xanthomonas gardneri*. Interestingly, all strains, with exception of *Novosphingobium* sp. PP1Y, belong to the family of the *Xanthomonadaceae*, which suggests that most of the biosynthetic gene clusters that could produce seven-ring lasso peptides are closely related to each other. This is further highlighted when comparing the precursor peptides of these six *Xanthomonadaceae* precursor peptides to each other (see Supporting Information Figure S1).



Xanthomonas gardneri ATCC 19865

Novosphingobium sp. PP1Y
Rhodanobacter sp. 115
Rhodanobacter sp. 116-2
Rhodanobacter sp. 2APBS1
Xanthomonas citri pv. *magniferaeindicae*

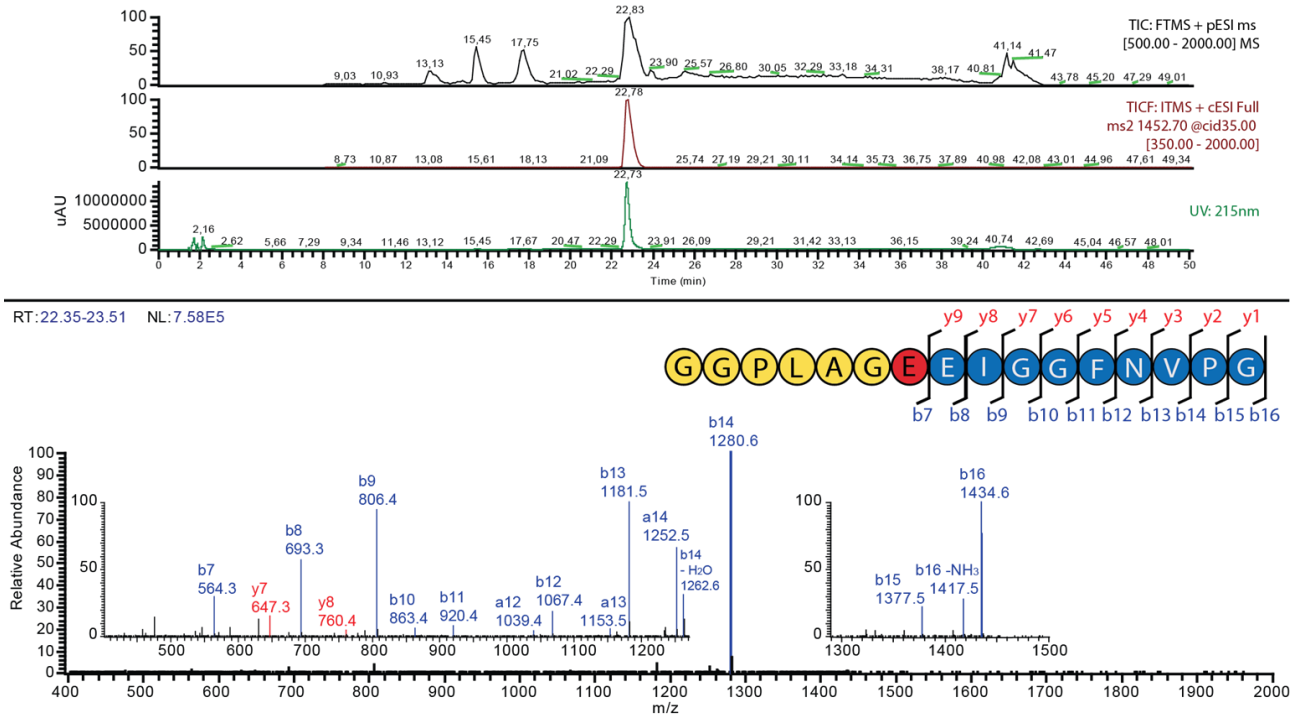
```
R115_A          MTQTQEIVMNTNENIRNNASEDVIVLGVASVETQ-GGPGSNEAVG-FGGPVIPGISEE
R116_A          MTRPQETEMNTNENIRNDTPEDVIVLGVASVETK-GPMGSTEPEGGLGRPVIPGISEE
R2APBS1_A             MNTNENIRTNAPEDVIELGIASVETK-GEFGVTEGGG-QGKPMIPGISEE
Xcit_A             MKTFVDSAPS-QEDHGKDIIVLGIASVETQ-GGAGAGEVNG-MS-P-IAGISEE
Xga_A1               MNS--NDTTHSDASNEITVLGVASTDTK-GGPLAGEEIGGFN---VPGISEE
Xga_A2               MDTSNNDARTTALDQDLIVLGVASLDTQ-GGPLAGEEMGGIT---TLGISQD
                                  :::  **:** :*: *     *  *        ***::

NPP1Y_A       MESRSRMEKIMEKTIDTLEDKVVDLGSVTEETR-GIVGALEDQQGGQRAALGLTDD
```

Supplementary Figure S1 - Comparison of all putative lasso peptide biosynthetic gene clusters that encode precursor peptides that could possibly be processed into lasso peptides with seven-residue macrolactam rings. The precursor peptides found in strains belonging to the *Xanthomonadaceae* family are shown in a direct sequence alignment, while the precursor peptide found in *Novosphingobium* sp. PP1Y is depicted below.

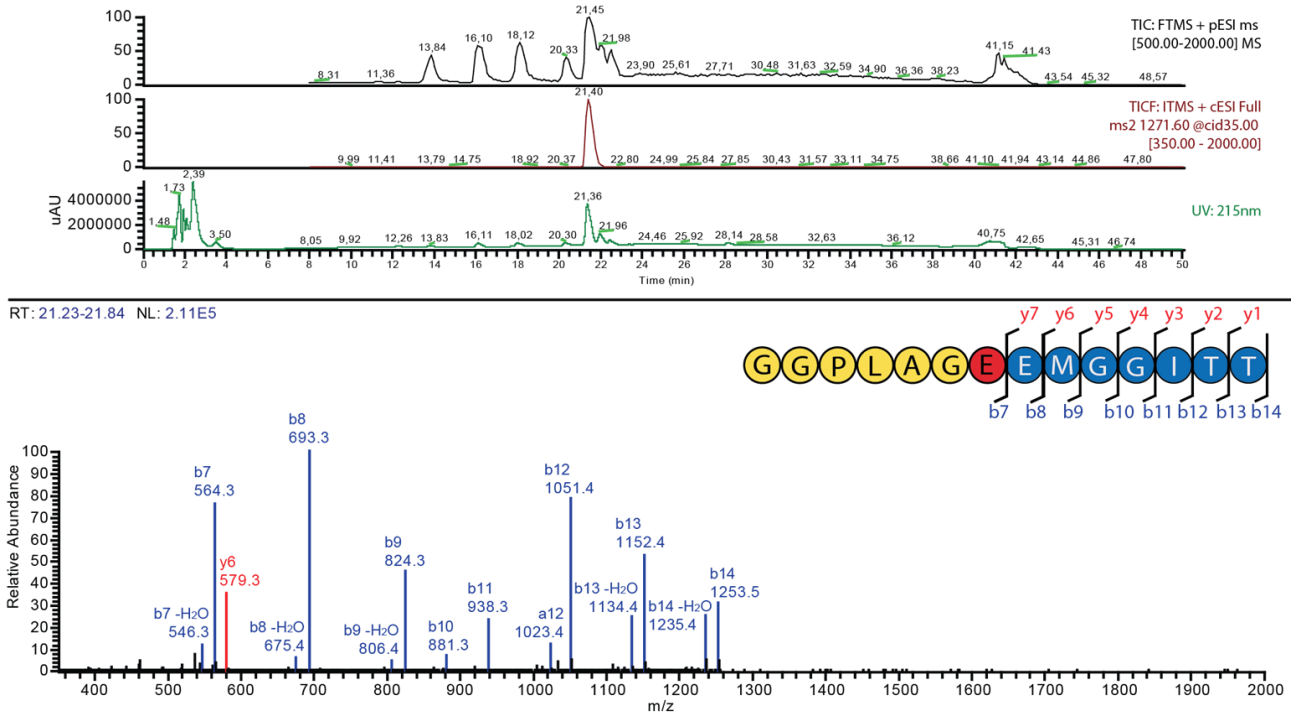
From all these proposed precursor peptides, four carry an additional amino acid that could be used for macrolactam ring formation, namely the precursor peptide of *Novosphingobium* sp. PP1Y (Asp8), the one of *Rhodanobacter* sp. 2APBS1 (Glu9) and both from *Xanthomonas gardneri* (Glu8 in both cases).

Tandem Mass Spectrometry of Xanthomonins I-III. Tandem Mass Spectrometry was used to confirm the primary structure of all xanthomonins and the presence and location of the proposed macrolactam rings.[14] The latter is possible, as a single peptide bond breakage in the macrolactam ring would just open the ring without an alteration of the overall mass. Thus, fragment ions of the macrolactam rings are only detectable in case of a double bond breakage. Since double bond breakages occur with only a low probability for such small peptides when using CID, the corresponding signals can normally not be detected. Because of this, all b- and y-fragment ions that are detectable with high intensities are generally associated with single bond breakages in the C-terminal tail region of the lasso peptide. Based on this, one can determine the presence and location of a macrolactam ring by searching for the smallest detectable b-fragment (and in some cases for the biggest detectable y-fragment). In the case of a lasso peptide with a seven-residue macrolactam ring, this would be the b7-fragment ion, which was detected for all xanthomonins, while a signal corresponding to a b6-fragment ion could not be found. The fragmentation spectra of all xanthomonins alongside the assignments of all high intensity signals to the corresponding fragment ions are shown in Supplementary Figures S2a-S2c.



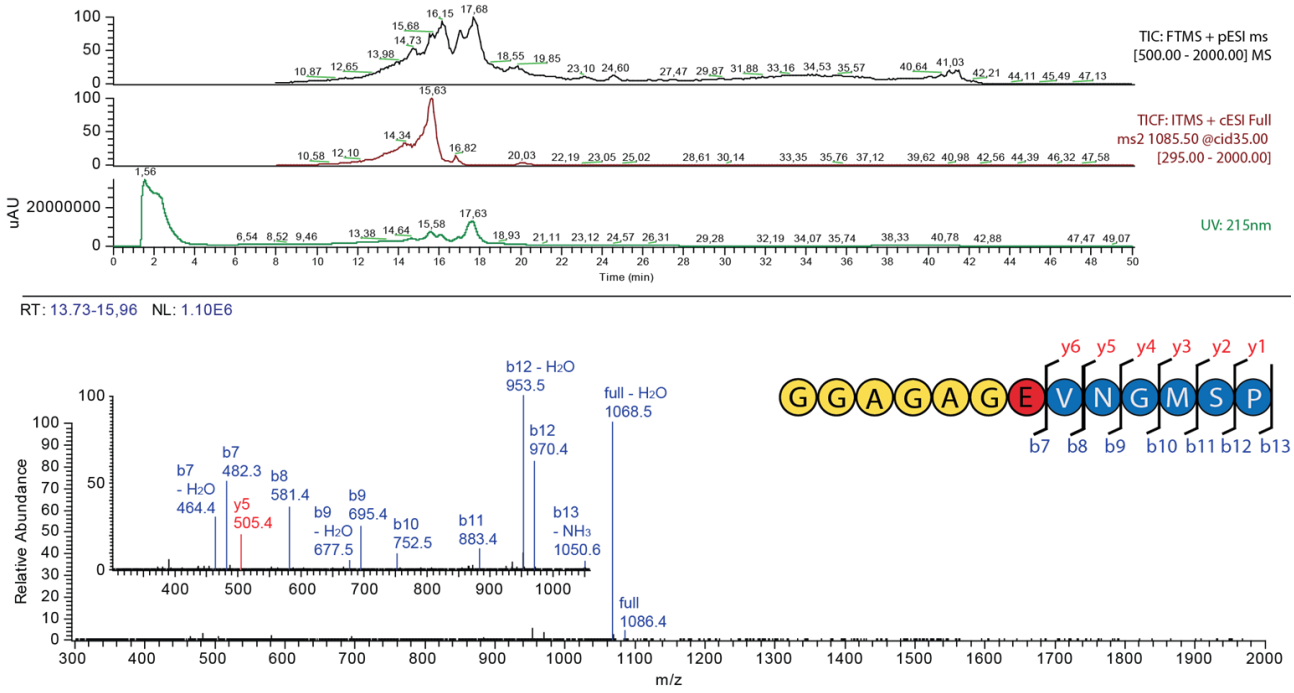
fragment ion	charge	predicted mass	observed mass	intensity	fragment ion	charge	predicted mass	observed mass	intensity
b16	+1	1434.7	1434.6	5343	b11	+1	920.4	920.4	2370
b16 - NH_3	+1	1417.7	1417.5	1476	b10	+1	863.4	863.4	1840
b15	+1	1377.7	1377.5	1166	b9	+1	806.4	806.4	28888
b15 - NH_3	+1	1360.7	1360.5	208	b9 - H_2O	+1	788.4	788.5	1270
b14	+1	1280.6	1280.5	758070	b8	+1	693.3	693.3	17409
b14 - H_2O	+1	1262.6	1262.6	9449	b8 - H_2O	+1	675.3	675.4	770
a14	+1	1252.6	1252.5	20182	b7	+1	564.3	564.3	9108
a14 - H_2O	+1	1234.6	1234.6	1265	b7 - H_2O	+1	546.3	546.4	2054
a14 - NH_3	+1	1235.6	1235.5	1923	a7	+1	536.3	536.3	1152
b13	+1	1181.6	1181.5	30563	y8	+1	760.4	760.4	1432
a13	+1	1153.6	1153.5	1622	y7	+1	647.3	647.3	4574
b12	+1	1067.5	1067.4	5703	z7	+1	630.3	630.4	3900
a12	+1	1039.5	1039.4	1191	y6	+1	590.3	590.3	124

Supplementary Figure S2a - Mass analysis of the -4 aa truncation of xanthomonin I. The total ion current (TIC) is shown on the top, followed by the extracted ion current (EIC) of the fragmented singly charged ion and the UV trace of this measurement. In the middle, the actual fragmentation spectra is depicted with an inset schematic of the primary structure of the -4 aa truncation of xanthomonin I. In the table below, all detected fragment ions with their charge, their predicted masses, their observed masses and their signal intensities are listed.



fragment ion	charge	predicted mass	observed mass	intensity	fragment ion	charge	predicted mass	observed mass	intensity
b14	+1	1253.6	1253.5	65631	b9	+1	824.4	824.3	95905
b14 - H_2O	+1	1235.6	1235.4	52972	b9 - H_2O	+1	806.4	806.4	9993
b14 - 2 H_2O	+1	1217.6	1217.4	2951	a9	+1	796.4	796.2	2126
a14	+1	1225.6	1225.4	1049	b8	+1	693.3	693.3	211200
a14 - H_2O	+1	1207.6	1207.6	1206	b8 - H_2O	+1	675.3	675.4	12914
b13	+1	1152.5	1152.4	111249	a8	+1	665.3	665.3	2180
b13 - H_2O	+1	1134.5	1134.4	51823	b7	+1	564.3	564.3	161058
b13 - 2 H_2O	+1	1116.5	1116.5	3042	b7 - H_2O	+1	546.3	546.3	25017
a13	+1	1124.5	1124.5	2306	a7	+1	536.3	536.3	15703
b12	+1	1051.5	1051.4	165655	a7 - H_2O	+1	518.3	518.4	5938
b12 - H_2O	+1	1033.5	1033.5	8923	y7	+1	708.3	708.3	1445
a12	+1	1023.5	1023.4	26092	y6	+1	579.3	579.3	74982
a12 - H_2O	+1	1005.5	1005.4	4293	y6 - H_2O	+1	561.3	561.3	4590
b11	+1	938.4	938.3	49018	y6 - 2 H_2O	+1	543.3	543.3	6377
b11 - H_2O	+1	920.4	920.5	3634	y5	+1	448.2	448.3	539
b10	+1	881.4	881.3	14419	y4	+1	391.2	391.1	2399
b10 - H_2O	+1	863.4	863.4	2600					

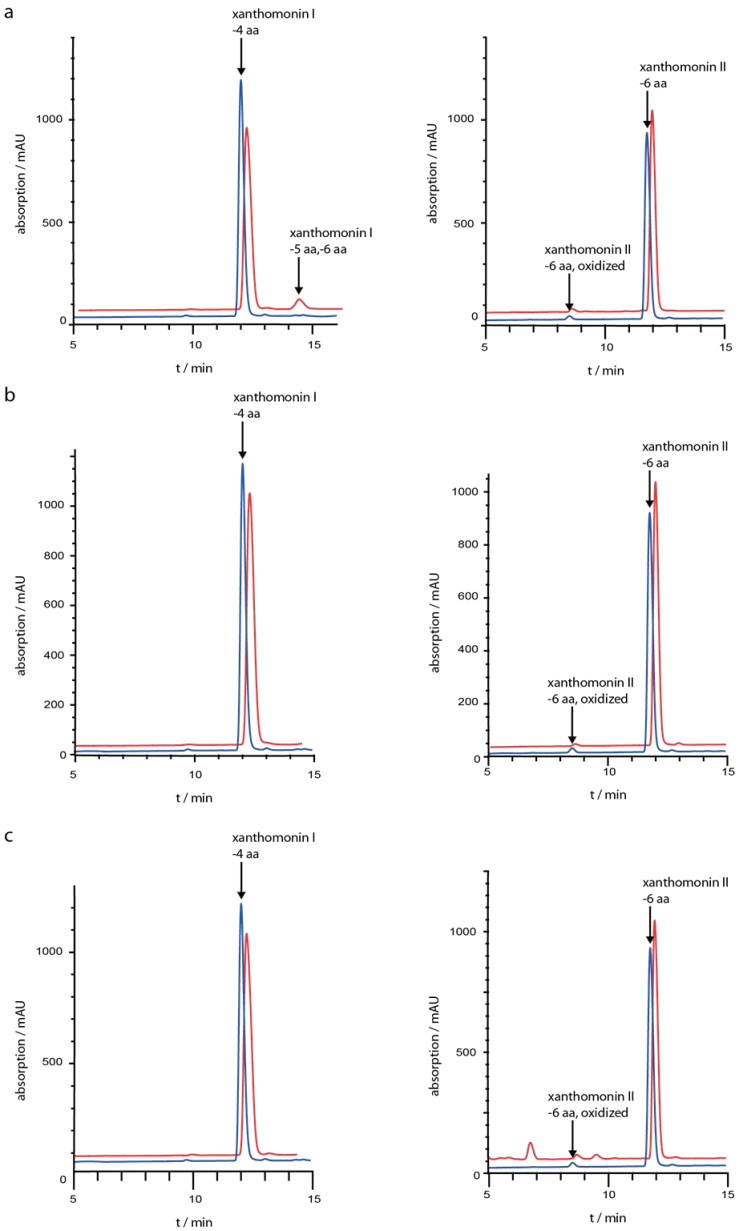
Supplementary Figure S2b - Mass analysis of the -6 aa truncation of xanthomonin II. The total ion current (TIC) is shown on the top, followed by the extracted ion current (EIC) of the fragmented singly charged ion and the UV trace of this measurement. In the middle, the actual fragmentation spectra is depicted with an inset schematic of the primary structure of the -6 aa truncation of xanthomonin II. In the table below, all detected fragment ions with their charge, their predicted masses, their observed masses and their signal intensities are listed.



fragment ion	charge	predicted mass	observed mass	intensity	fragment ion	charge	predicted mass	observed mass	intensity
full	+1	1086.2	1086.4	46905	b9	+1	695.3	695.4	13067
full - H_2O	+1	1068.2	1068.5	1098662	b9 - H_2O	+1	677.3	677.5	2637
b13 - NH_3	+1	1050.5	1050.6	2330	b8	+1	581.3	581.4	18863
b12	+1	970.4	970.4	32987	b7	+1	482.2	482.3	26848
b12 - NH_3	+1	953.4	953.5	52945	b7 - H_2O	+1	464.2	464.4	15836
b12 - H_2O	+1	952.4	952.6	4953	y5	+1	505.2	505.4	10601
b11	+1	883.4	883.4	6115	y5 - H_2O	+1	487.2	487.4	1028
b10	+1	752.3	752.5	4620					

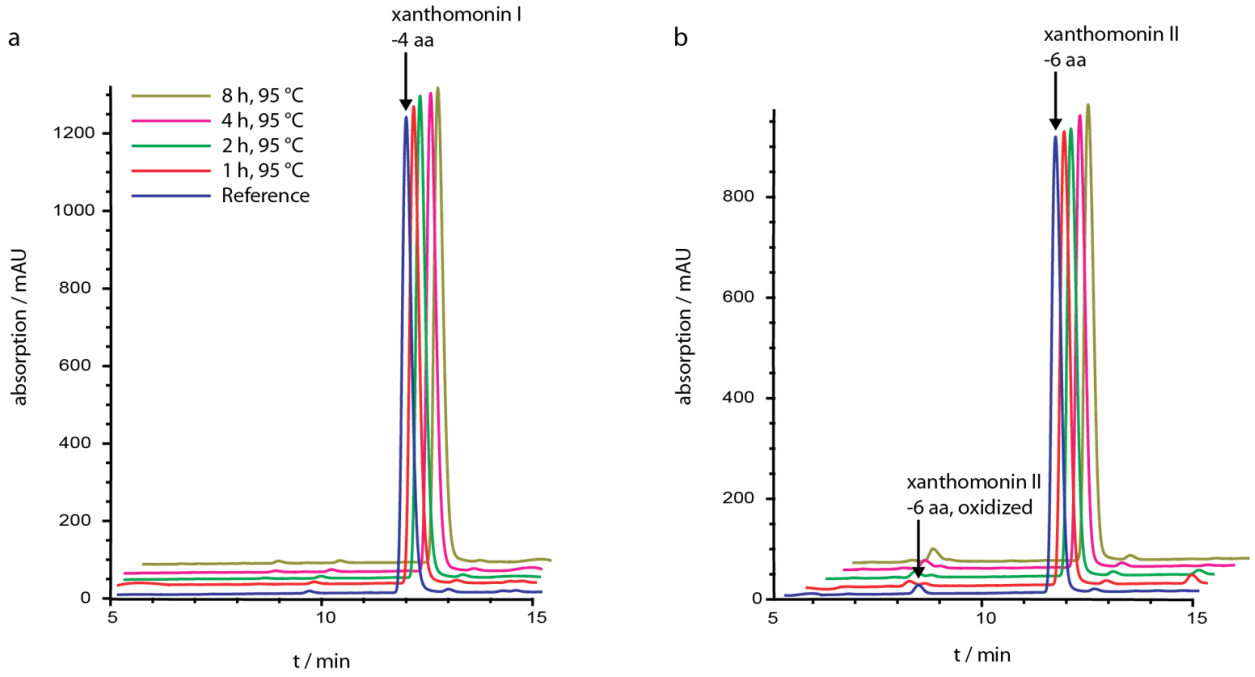
Supplementary Figure S2c - Mass analysis of the -7 aa truncation of xanthomonin III. The total ion current (TIC) is shown on the top, followed by the extracted ion current (EIC) of the fragmented singly charged ion and the UV trace of this measurement. In the middle, the actual fragmentation spectra is depicted with an inset schematic of the primary structure of the -7 aa truncation of xanthomonin III. In the table below, all detected fragment ions with their charge, their predicted masses, their observed masses and their signal intensities are listed.

Protease Stability Assays for Xanthomonins I and II. Protease stability of xanthomonins I and II was assessed against carboxypeptidase Y, chymotrypsin and proteinase K and analyzed via LC-MS. In agreement with previous studies, both lasso peptides showed a high stability against proteolytic degradation.[1, 14-15] Only carboxypeptidase Y treatment of xanthomonin I resulted in the slow formation of trace amounts of the -5 and -6 aa truncation, of which the latter has the same length as the main product of the heterologous production of xanthomonin II. The recorded chromatograms of these assays are shown in Supplementary Figure S3.



Supplementary Figure S3 - Protease stability assays of the -4 aa truncation of xanthomonin I (left side) and the -6 aa truncation of xanthomonin II (right side) against (a) carboxypeptidase Y, (b) chymotrypsin and (c) proteinase K. The UV traces of the untreated references are shown in blue, while the UV traces of the samples after protease treatment are shown in red.

Thermal Stability Assays for Xanthomonin I and II. To test the thermal stability of xanthomonins I and II, both lasso peptides were incubated for 1 h, 2 h, 4 h or 8 h at 95 °C and then compared to an unheated reference by LC-MS. In both cases, the thermal treatment showed no effect even after 8 h at 95 °C. This highlights the high thermal stability of lasso peptides containing a seven-residue macrolactam ring. The recorded chromatograms of these assays are shown in Supplementary Figure S4.



Supplementary Figure S4 - Thermal stability assays of (a) the -4 aa truncation of xanthomonin I (left side) and (b) the -6 aa truncation of xanthomonin II. In both cases, no change in molecular mass or retention time could be observed, even after incubation for 8 h at 95 °C.

Combined Carboxypeptidase Y and Thermal Stability Assays for Xanthomonin II Variants. As shown in previous studies,[1, 13-14] carboxypeptidase Y treatment can be used to differentiate between a lasso and a branched cyclic peptide. When combined with a thermal denaturation experiment, carboxypeptidase Y can also be used to determine if the lasso peptide unthreads during incubation at 95 °C. Such assays were performed for the plug amino acid scan variants I12X/T13A/T14A/L15A of xanthomonin II and a wild type control (see Supplementary Tables S6a and S6b).

For the wild type, carboxypeptidase Y digestion yielded the same results independent if the sample was incubated at 95 °C for 4 h or not. In both cases, the digestion could not produce any shorter main products than the -6 aa truncation of xanthomonin II. Similar results were obtained for the variants carrying an Ile, Leu, Thr or Val at position 12. For all these variants, the by far major product of the carboxypeptidase Y digestion was the -6 aa truncation, with only small traces of other truncations detectable. This data suggests that all these variants are indeed heat stable lasso peptides.

In contrast, both variants carrying either Ala or Ser at position 12 yielded the -12 aa truncation as main products after the carboxypeptidase Y treatment. The fact that this was observed for both the thermally treated and untreated samples, suggests that they already had a branched cyclic topology to begin with. In case of the Ala variant this is not surprising, as all tail amino acids from positions 12 to 15 were replaced with Ala and this should allow an easy unthreading of the lasso peptide directly after its maturation. This is also reflected by the fact that after isolation, the -10 aa and -12 aa truncations were the only truncations that had an detectable UV peak. Interestingly, even though the variant carrying a Ser residue at position 12 behaves similar when treated with carboxypeptidase Y, the main products after extraction are not only a -10 aa, but also a -4 aa truncation. Additionally, the overall amounts isolated of this variant were around twice as much when compared to the Ala variant. This may be a hint that with a Ser in the role of the plug amino acid, the lasso fold does not unthread as easily as it is the case with Ala. Hence, this variant could not be degraded as quickly *in vivo*, which is why significant amounts of the -4 aa truncation could still be isolated. Nonetheless, the carboxypeptidase Y assay shows that after isolation all truncations of the Ser variant were present in the branched cyclic topology, which leaves the question unresolved, if the compound was already completely unthreaded before isolation or if the unthreading occurred during the isolation procedure. Nevertheless, the result shows that Ser is at best a very poor plug amino acid for xanthomonin II and in summary, these studies reveal that all amino acids bigger than Ser are suitable to maintain a heat stable lasso fold in the xanthomonin II scaffold.

Supplementary Table S6a. Results of the combined heat stability and carboxypeptidase Y assays of the extracts of the I12X T13A T14A L15A variants of xanthomonin II. The major signals of each measurement are highlighted in bold.

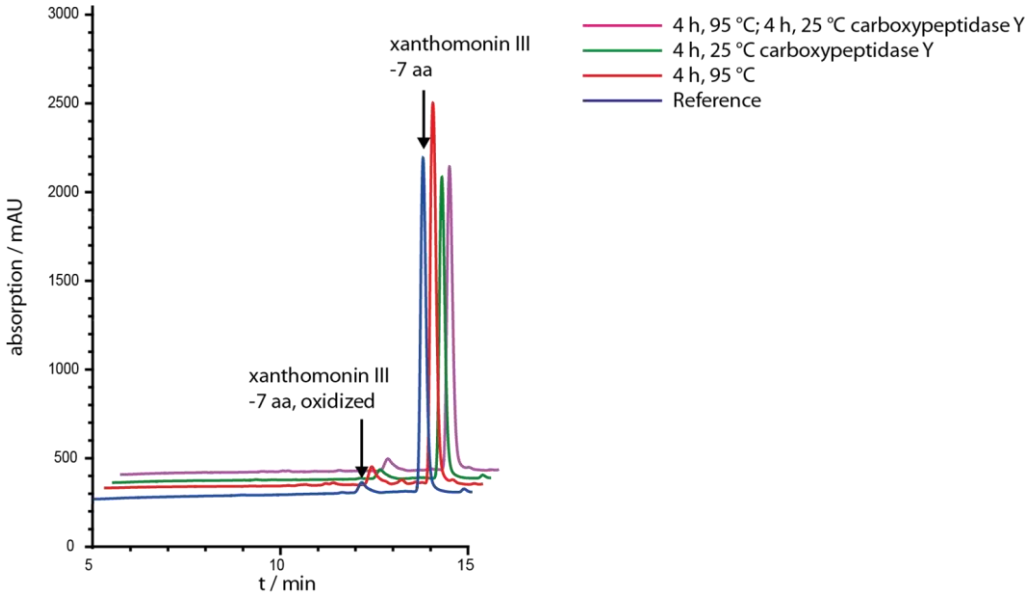
name	condition	UV peak areas and retention time of all detectable compounds
wild type	reference	**1.8E+7 (22.1 min, -4 aa), 8.3E+6 (22.7 min, -5 aa), 6.2E+7 (21.6 min, -6 aa)**
	4 h, 95 °C	**1.5E+7 (22.6 min, -4 aa), 6.2E+6 (23.2 min, -5 aa), 4.6E+7 (21.7 min, -6 aa)**
	carboxypeptidase Y	**5.8E+7 (21.4 min, -6 aa)**
	4 h, 95 °C; carboxypeptidase Y	**6.0E+7 (21.5 min, -6 aa)**
I12A T13A T14A L15A	reference	- (19.4 min, full), - (17.6 min, -4 aa), – (17.6 min, -6 aa), - (17.4 min, -7 aa), - (17.5 min, -8 aa), - (16.7 min, -9 aa), **3.8E+6 (17.0 min, -10 aa)**, - (17.7 min, -11 aa), **6.4E+5 (14.3 min, -12 aa)**, - (13.9 min, -13 aa)
	4 h, 95 °C	- (19.4 min, full), - (17.6 min, -4 aa) – (17.7 min, -6 aa), - (17.5 min, -7 aa), - (17.5 min, -8 aa), - (16.7 min, -9 aa), **4.1E+6 (17.0 min, -10 aa)**, - (17.8 min, -11 aa), **6.3E+5 (14.2 min, -12 aa)**, - (13.9 min, -13 aa)
	carboxypeptidase Y	**1.9E+6 (14.3 min, -12 aa)**, - (14.1 min, -13 aa)
	4 h, 95 °C; carboxypeptidase Y	**1.8E+6 (14.4 min, -12 aa)**, - (14.0 min, -13 aa)
T13A T14A L15A	reference	**5.1E+7 (20.8 min, -4 aa), 3.5E+6 (21.1 min, -5 aa), 9.9E+6 (21.7 min, -6 aa),** - (14.2 min, -12 aa)
	4 h, 95 °C	**5.6E+7 (20.9 min, -4 aa), 3.2E+6 (21.2 min, -5 aa), 9.8E+6 (21.8 min, -6 aa)**, - (14.2 min, -12 aa)
	carboxypeptidase Y	- (21.0 min, -4 aa), **6.0E+7 (22.0 min, -6 aa)**, - (21.9 min, -7 aa), - (14.3 min, -12 aa)
	4 h, 95 °C; carboxypeptidase Y	**6.3E+7 (21.9 min, -6 aa)**, - (21.8 min, -7 aa), - (14.2 min, -12 aa)
I12L T13A T14A L15A	reference	- (21.7 min, full), - (23.0 min, -3 aa), **5.9E+7 (21.3 min, -4 aa), 5.6E+6 (21.6 min, -5 aa), 3.4E+6 (22.1 min, -6 aa)**
	4 h, 95 °C	- (21.6 min, full), - (22.8 min, -3 aa), **5.5E+7 (21.2 min, -4 aa), 7.0E+6 (21.6 min, -5 aa), 5.1E+6 (22.1 min, -6 aa)**
	carboxypeptidase Y	- (21.3 min, -4 aa), **5.7E+7 (22.2 min, -6 aa)**, - (22.0 min, -7 aa), - (14.2 min, -12 aa)
	4 h, 95 °C; carboxypeptidase Y	**5.6E+7 (22.1 min, -6 aa)**, - (21.9 min, -7 aa), - (14.3 min, -12 aa)
I12V T13A T14A L15A	reference	- (20.6 min, full), **3.8E+7 (19.9 min, -4 aa), 4.2E+6 (20.4 min, -5 aa), 9.9E+6 (20.9 min, -6 aa)**
	4 h, 95 °C	- (20.6 min, full), **4.1E+7 (19.9 min, -4 aa), 4.8E+6 (20.3 min, -5 aa), 1.1E+7 (20.9 min, -6 aa)**, - (14.2 min, -12 aa)
	carboxypeptidase Y	**4.4E+7 (20.9 min, -6 aa)**, - (14.2 min, -12 aa)
	4 h, 95 °C; carboxypeptidase Y	**4.5E+7 (21.1 min, -6 aa)**, - (14.2 min, -12 aa)

Supplementary Table S6b. Results of the combined heat stability and carboxypeptidase Y assays of the extracts of the I12X T13A T14A L15A variants of xanthomonin II. The major signal(s) of each measurement are highlighted in bold.

name	condition	UV peak areas and retention time of all detectable compounds
I12S T13A T14A L15A[a]	reference	- (19.2 min, full), **4.4E+6 (17.2 min, -4 aa/-10 aa)**, - (17.3 min, -6 aa), **2.7E+5 (16.5 min, -7 aa)**, - (17.0 min, -9 aa), - (18.2 min, -11 aa), - (14.3 min, -12 aa), - (14.0 min, -13 aa)
	4 h, 95 °C	- (19.2 min, full), **4.2E+6 (17.2 min, -4 aa/-10 aa)**, **2.0E+5 (16.5 min, -7 aa)**, - (17.0 min, -9 aa), - (18.0 min, -11 aa), - (14.3 min, -12 aa), - (14.0 min, -13 aa)
	carboxypeptidase Y	**2.1E+6 (14.5 min, -12 aa)**, - (14.0 min, -13 aa)
	4 h, 95 °C; carboxypeptidase Y	**2.3E+6 (14.3 min, -12 aa)**, - (14.0 min, -13 aa)
I12T T13A T14A L15A	reference	**5.8E+7 (18.4 min, -4 aa), 4.1E+5 (18.9 min, -5 aa), 4.3E+6 (19.6 min, -6 aa)**
	4 h, 95 °C	**5.2E+7 (18.2 min, -4 aa), 2.0E+5 (18.9 min, -5 aa), 4.3E+6 (19.8 min, -6 aa),** - (16.9 min, -9 aa), - (17.1 min, -10 aa), - (14.2 min, -12 aa), - (14.0 min, -13 aa)
	carboxypeptidase Y	**6.1E+7 (19.6 min, -6 aa), 6.9E+5 (18.6 min, -7 aa)**, - (14.2 min, -12 aa)
	4 h, 95 °C; carboxypeptidase Y	- (18.9 min, -4 aa), **5.8E+7 (19.7 min, -6 aa), 6.9E+5 (18.8 min, -7 aa)**, - (14.4 min, -12 aa), - (14.0 min, -13 aa)

[a] UV signals of -4 aa and -10 aa truncations strongly overlap. The intensity of their MS signals suggests an approximately equimolar ratio.

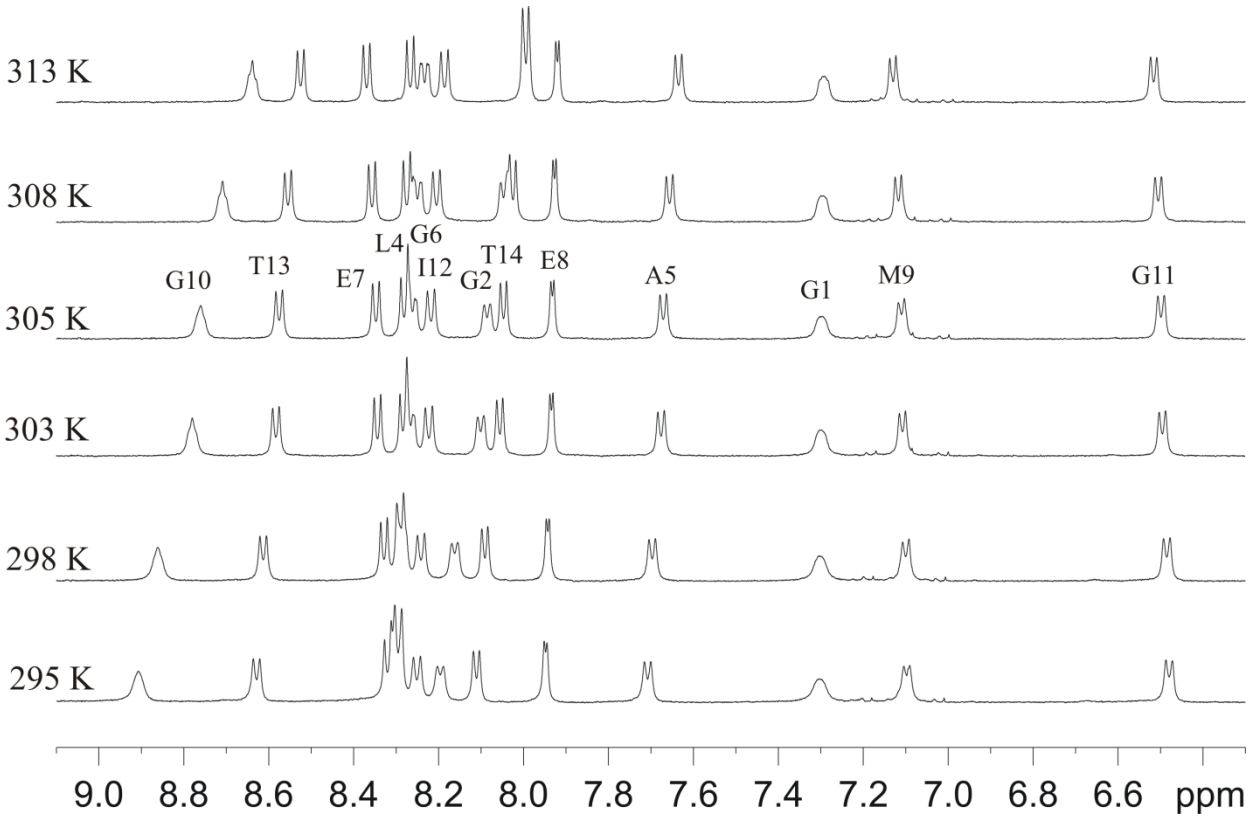
Combined Carboxypeptidase Y and Thermal Stability Assays for Xanthomonin III. Combined carboxypetidase Y and thermal stability assays were performed with a purified sample of the -7 aa truncation of xanthomonin III. These assays revealed that xanthomonin III is resistant to both carboxypeptidase Y digestion and thermal denaturation, proving it to be as stable as xanthomonins I and II. Furthermore, these results, in combination with its heterologous origin and the data obtained from the tandem mass spectrometry analysis, give strong evidence of xanthomonin III being a true lasso peptide, even though its 3D structure was not elucidated. The recorded chromatograms of these assays are shown in Supplementary Figure S5.



Supplementary Figure S5 - Combined carboxypeptpidase Y and thermal stability assays of the -7 aa truncation of xanthomonin III.

Analysis of the Xanthomonin III Variants Suggests that Met11 Acts as Plug Amino Acid. The generated and analysed variants of xanthomonin III provide two lines of evidence of Met11 acting as plug amino acid. This first indirect evidence is the lack of production of the M11A/S12A/P13A/I14A variant that suggests that this quadruple alanine mutation removes all tail amino acids that could act as plug for xanthomonin III, starting with Met11. The second evidence is provided by the fact that all other variants, namely S12A/P13A/I14A, P13A/I14A and I14A, are produced in comparable amounts to the wild type xanthomonin III. This suggests that these three amino acids are not needed for maintaining the lasso fold as long as Met11 is present, further emphasizing its proposed role as plug amino acid. Furthermore, the observation that all produced variants share the same truncation pattern with the wild type xanthomonin III, is an indication that all are present in the same lasso fold, limiting their truncation to points that are equidistant to the position where the tail threads through the macrolactam ring. Taking all this information together, this data is solid body of evidence for xanthomonin III being a true lasso peptide that uses Met11 as plug amino acid.

Structure Elucidation of Xanthomonin II Using NMR Spectroscopy. The primary structure of the -6 aa truncation of xanthomonin II consists of 14 amino acids. The residue at position 1 is a Gly, while both amino acids at positions 7 and 8 are Glu. Therefore, the question arises if Glu7 or Glu8 is involved in the isopeptide bond formation needed for the macrocyclization. Both tandem mass spectrometry and mutational analysis suggest Glu7 to be the ring forming amino acid. Therefore, xanthomonin II represents not only one of the shortest lasso peptides known so far, but also features the smallest ring size for a lasso peptide yet. This reduces the minimal proven ring size to be seven amino acids and may be the minimum possible for lasso peptides. Determining its 3D structure with great care is thus of basic importance.
The ^{1}H spectra in DMSO-d_6 at temperatures between 295 and 313 K with 5 K increments are shown in Supplementary Figure S6.



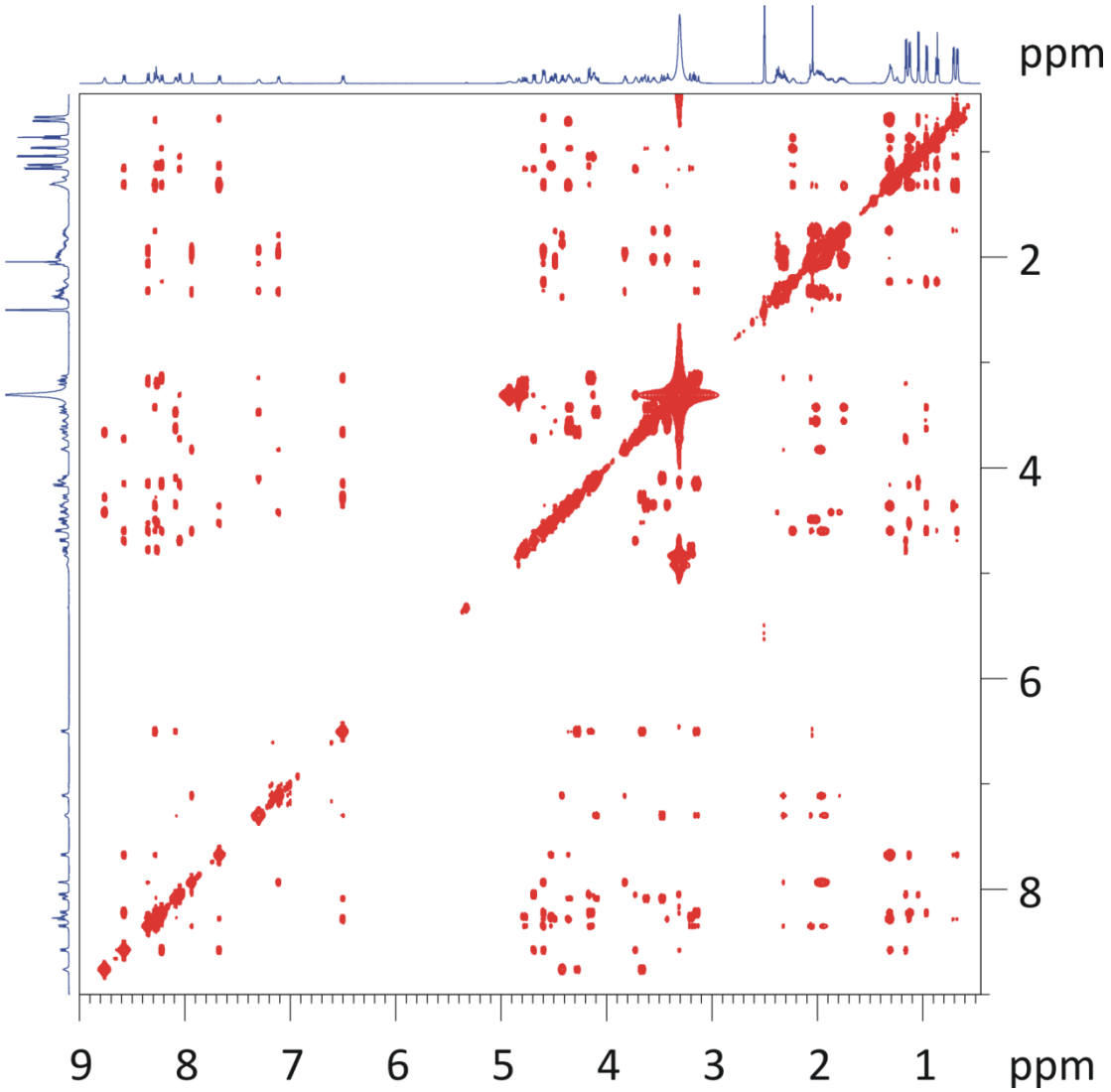
Supplementary Figure S6 - Variable temperature ^{1}H spectra in NH region of the -6 aa truncation of xanthomonin II in DMSO-d_6 at 8.8 mM.

For clarity, only the region of 9.10 to 6.30 ppm is shown in the spectra. Labels for the signal assignments of the amide protons are shown. Neat and well resolved single sets of signals were observed. A diverse distribution of the temperature response of these signals revealed a stable lasso fold of this compound. The best signal resolution was observed at 305 K and a full signal assignment (see Supplementary Table S7) was thus carried out at this temperature.

Supplementary Table S7. Assignment of ^{1}H signals (in ppm) of the -6 aa truncation of xanthomonin II (14 amino acids in length) in DMSO-d_6 at 305 K.

AA	NH	αH	βH	others
Gly1	7.2984	4.0945; 3.4672	/	/
Gly2	8.0857	4.3472; 3.6177	/	/
Pro3		4.487	2.010	γCH_2: 1.748; 2.003 δCH_2: 3.546; 3.416
Leu4	8.2800	4.3502	1.2942	γCH: 1.323 δCH_3: 0.7061; 0.6702
Ala5	7.6705	4.5084	1.1242	/
Gly6	8.2612	4.7757; 3.193	/	/
Glu7	8.3473	4.5977	1.936	γCH2: 2.317; 2.052
Glu8	7.9315	3.8321	1.9747	γCH2: 2.333
Met9	7.1098	4.4187	1.857; 1.785	γCH2:2.373 εCH3:2.0433
Gly10	8.759	4.2741; 3.660	/	/
Gly11	6.4984	4.1449; 3.1413	/	/
Ile12	8.2175	4.5916	2.2289	γCH3: 0.9609 γCH2: 1.3180; 1.1330 δCH_3: 1.1574
Thr13	8.5749	4.6856	3.7190	γCH3: 1.1574
Thr14	8.0469	4.1649	4.1217	γCH3: 1.0391

All assignments of the ^{1}H signals were obtained by standard procedures.[11] A combination of DQF-COSY and NOESY produced sequential assignments (i.e. all αH and NH and their sequences in the backbone), and a combination of DQF-COSY and TOCSY allowed the determination of the side chains. One pure conformation was observed. Full assignment of ^{1}H signals was thus obtained (NOESY spectrum: Supplementary Figure S7, ^{1}H chemical shifts: Supplementary Table S7).



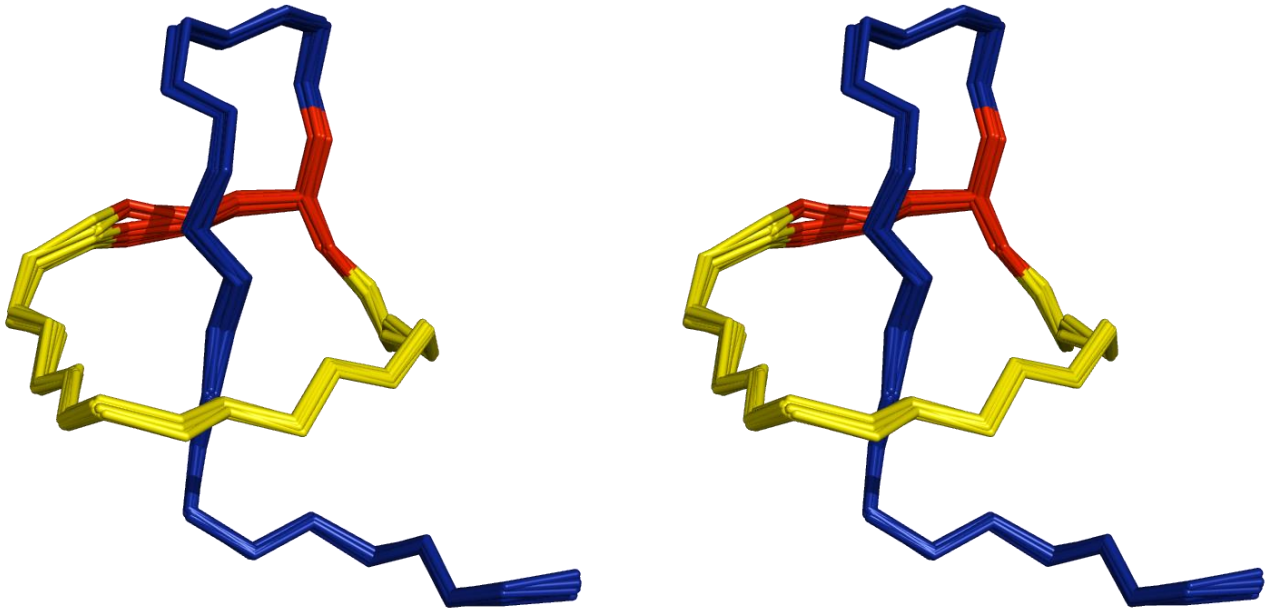
Supplementary Figure S7 - NOESY spectrum of the -6 aa truncation of xanthomonin II in DMSO-d_6 at 305 K with a mixing time of 300 ms.

Strong NOE contacts between the NH of Gly1 and both the βH and γH of Glu7 were observed, showing an internal linkage between these two residues. Strong sequential αH – δH NOESY cross-peaks between Gly2 and Pro3 were detected, which showed the trans-conformation of Pro3. The ^{1}H spectra between 295 and 313 K (Supplemtary Figure S6) show a low temperature coefficient of the NH of Gly1, Leu4, Ala5, Glu7, Glu8, Met9, Gly11, Leu12, and Thr13. A large number of long-range NOE contacts were observed. These are the $d_{NN}(i,j)$ connectivity between Gly1 – Gly11, Gly2 – Gly11, Leu4 – Gly11, and Ala5 – Thr13; the $d_{\alpha N}(i,j)$ connectivity between Glu7 – Ile12, Gly11 – Glu7, and Thr14 – Ala5; the $d_{\beta N}(i,j)$ connectivity between Leu4 – Thr13 and Glu7 – Gly1; and the $d_{\alpha\alpha}(i,j)$ connectivity between Ala5 – Gly10. Furthermore, long-range contacts between residue side chains were observed for Pro3 – Ile12, Leu4 – Ile12, and Leu4 – Thr13. All this experimental evidence revealed a very compact 3D structure and is in favor of a lasso structure for xanthomonin II.

Structure calculations were performed with the program CYANA 2.1.[13] The internal linkage was realized

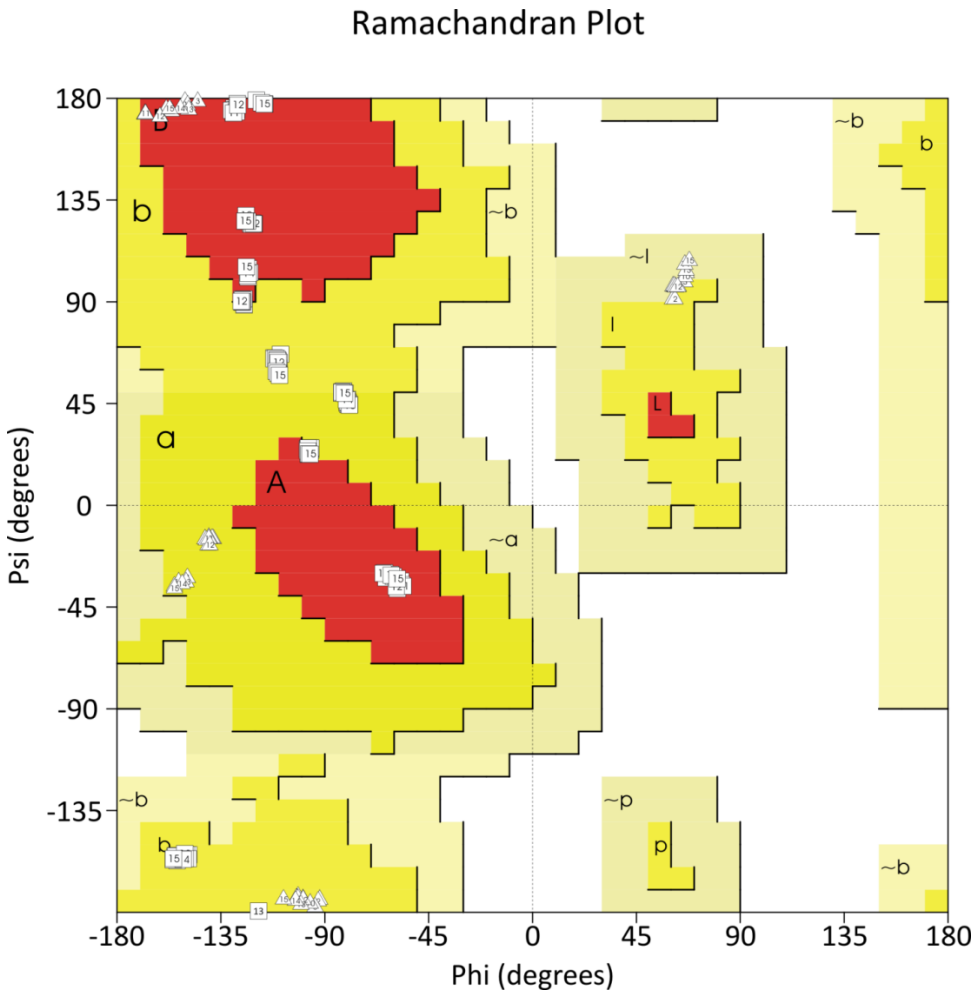
by setting the distance constraints between N of Gly1 and Cδ of Glu7 to be 1.33 Å. NOE cross-peaks observed in the 100 ms mixing time NOESY experiment were manually converted into distance constraints. In this way, 97 unambiguous distance constraints were obtained. 37 for the backbone, 25 for long-range interactions, and 35 for the side chains. Thus, there was an average of 7.5 distance constraints per residue. In addition, constraints of torsion angles φ and χ^1 were determined by analyzing the vicinal coupling constants $^3J_{HN\alpha}$ and $^3J_{\alpha\alpha}$. Only unambiguous coupling constants were used. Thus, $^3J_{HN\alpha} \geq 9$ Hz was observed for Gly2, Leu4, Ala5, Gly6, Glu7, Gly11, Ile12, and Thr13. The torsion angles φ of these residues were restrained to -120° ± 30°. For the residues Gly1, Glu8, Met9, and Gly10, $^3J_{HN\alpha} < 9$ Hz was detected. Thus their torsion angles φ were restrained to -70° ± 30°. Stereospecific assignment of the prochiral methylene protons was done for Gly1, Gly2, Pro3, Gly6, and Gly11. The side chain conformations of Met (t^2g^3), Thr13 (g^2g^3), and Thr14 (t^2g^3) were determined. For $t^2g^3$and g^2g^3 conformations around the Cα-Cβ bond the torsion angle χ^1 was constrained in the range of -60 ± 30°and 60 ± 30°, respectively.

The above mentioned constraints were used in the simulated annealing protocol for the calculation with the CYANA 2.1 program. The calculation was initiated with 50 random conformers and the resulting structures were processed with the program package Sybyl 7.3[16] to include the covalent linkage between the nitrogen of Gly1 and Cδ of Glu7, followed by energy minimization under NMR constraints using TRIPOS force field within Sybyl. Thus, on the basis of low energies and minimal violations of the experimental data, a family of 15 structures was chosen. These 15 energy-minimized conformers show an average root-mean-square deviation (rmsd) of 0.02 Å and were chosen to represent the solution structure of the -6 aa truncation of xanthomonin II (Supplementary Figure S8).



Supplementary Figure S8 - Crossed-eyes stereoview of the superimposition of the 15 lowest energy structures of the -6 aa truncation of xanthomonin II (PDB code 2MFV). Structures are superimposed over all backbone atoms. The isopeptide bond forming Glu7 is highlighted in red, while the remaining ring amino acids are colored in yellow and the tail amino acids are shown in blue.

The superimposition of the 15 lowest structures that are shown in Supplementary Figure S8 represents the lasso fold of xanthomonin II in DMSO-d_6 at 32 ° C. This is verified by a quality comparison of the 15 minimum energy structures and their Ramachandran plots (Supplementary Figure S9). The structural statistics for the 15 lowest energy structures are shown in Supplementary Table S8.



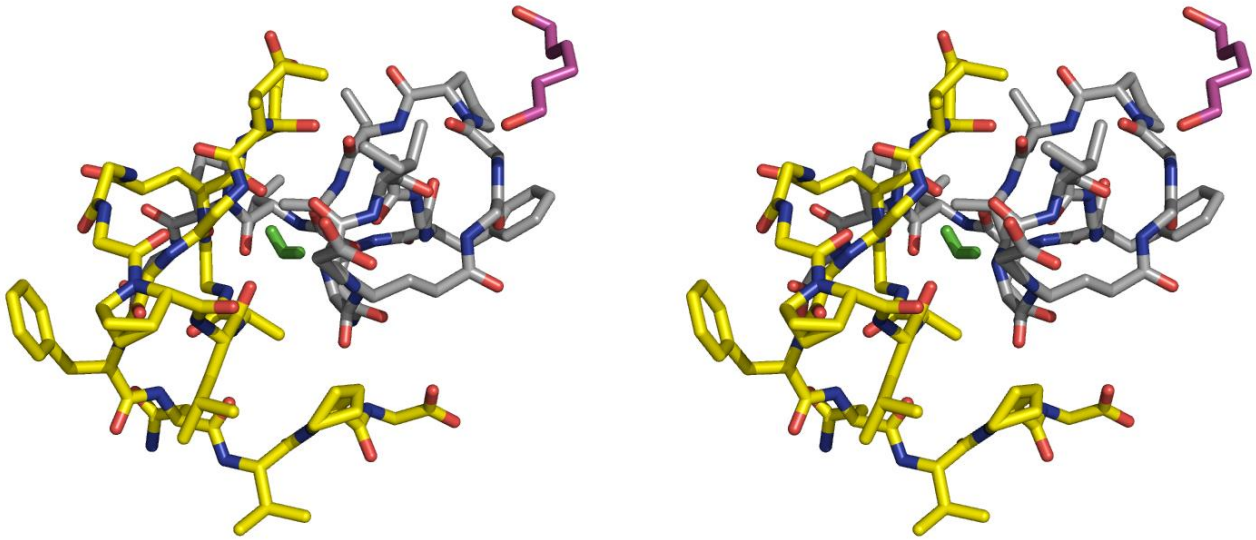
Supplementary Figure S9 - Ramachandran plot of the 15 lowest energy structures of the -6 aa truncation of xanthomonin II created by PROCHECK NMR.

Supplementary Table S8. Structural statistics for the 15 lowest energy structures selected to represent the structure of the -6 aa truncation of xanthomonin II in DMSO-d_6 at 32 °C.

restraining constraints	constraints violations
total: 120	distance violations, >0.5 Å: 0
distance, i=j: 30	RMS deviations: 0.02 Å
distance, |i-j|=1: 39	dihedral violations, > 5°: 0
distance, |i-j|>1: 28	RMS deviation: 2.8°
dihedral: 15	average pairwise RMS deviation (Gly^2-Thr^{13})
hydrogen bonds: 0	backbone atoms: 0.03 Å
constraints/residue: 8.6	all heavy atoms: 0.35 Å

Crystal Structure Elucidation of Xanthomonin I.

Data collection and crystal structure elucidation were done as described in the Experimental Procedures. Xanthomonin I crystallizes as a dimer with one bridging water atom in between both monomeric units. Addtionally, the structure contains 18 solvent water molecules and one molecule 1,6-hexanediol. The dimeric crystal structure is shown in Supplementary Figure S10. Data collection and refinement statistics are shown in Supplementary Table S9. Coordinates, bond lengths, angles and displacement parameters are shown in Supplementary Tables S10-S15.



Supplementary Figure S10 - Crossed-eyes stereoview of the crystal structure of xanthomonin I (PDB code 4NAG). The bridging water molecule is colored in green, while the 1,6-hexanediol and the xanthomin I molecules are colored by elements (nitrogen in blue, oxygen in red, carbon in magenta for the 1,6-hexanediol, in yellow for the first xanthomin I monomer and in gray for the second xanthomonin I monomer). The 18 solvent water molecules are not shown.

Even though the structures of xanthomonin I and II were elucidated by different means, both appear to be very similar. This is not surprising, as their ring sequence is identical and the only major difference in their amino acid composition is found in the tail region below the ring.

Supplementary Table S9. Crystal data and structure refinement for the -4 aa truncation of xanthomonin I.

Identification code	Xanthomonin I
Empirical formula	$C_{136}\ H_{244}\ N_{34}\ O_{63}$ $2(C_{65}H_{97}N_{17}O_{21}),C_6H_{12}O_2,19(H2O)'$
Formula weight	3363.64 g/mol
Temperature	100(2) K
Wavelength	0.8 Å
Crystal system	Monoclinic
Space group	P 2_1
Unit cell dimensions	a = 13.804(3) Å, a= 90°. b = 45.096(9) Å, b= 91.42(3)°. c = 13.890(3) Å, g = 90°.
Volume	8644(3) $Å^3$
Z	2
Density (calculated)	1.292 mg/m^3
Absorption coefficient	0.102 mm^{-1}
F(000)	3604
Theta range for data collection	0.903 to 25.935°.
Index ranges	-15<=h<=16, -54<=k<=54, -16<=l<=17
Reflections collected	42090
Independent reflections	24760 [R(int) = 0.0394]
Completeness to theta = 25.242°	84.6%
Refinement method	Full-matrix least-squares on F^2
Data / restraints / parameters	24760 / 111 / 2162
Goodness-of-fit on F2	1045
Final R indices [I>2sigma(I)]	R1 = 0.0535, wR2 = 0.1470
R indices (all data)	R1 = 0.0544, wR2 = 0.1484
Absolute structure parameter	-0.1(3)
Extinction coefficient	n/a
Largest diff. peak and hole	0.508 and -0.321 $e.Å^{-3}$

Supplementary Table S10a. Atomic coordinates (x 10^4) and equivalent isotropic displacement parameters (Å^{2}x 10^3) for -4aa truncation of xanthomonin I. U(eq) is defined as one third of the trace of the orthogonalized U^{ij} tensor.

	x	y	z	U(eq)		x	y	z	U(eq)
O(100)	9329(2)	6480(1)	6063(2)	32(1)	C22	7121(3)	8145(1)	9750(3)	33(1)
OH1	4722(3)	7698(1)	9979(3)	63(1)	O13	6943(4)	8477(1)	5672(3)	69(1)
C(100)	4299(6)	7906(2)	10577(6)	85(2)	N13	7162(3)	8449(1)	8314(3)	34(1)
C(101)	4237(9)	8209(2)	10036(8)	124(4)	C13	7269(4)	8410(1)	6481(4)	38(1)
C(102)	3941(12)	8465(2)	10658(9)	133(4)	C23	6814(4)	8549(1)	7353(4)	40(1)
C(103)	3973(14)	8758(3)	10083(17)	203(8)	C33	7024(6)	8881(1)	7393(5)	65(2)
C(104)	4849(19)	8925(7)	9650(30)	342(15)	C43	7880(7)	8908(2)	8060(6)	75(2)
C(105)	4760(30)	9230(6)	9145(13)	346(16)	C53	7714(4)	8681(1)	8831(4)	45(1)
OH2	4593(6)	9154(2)	8129(10)	151(4)	O14	7746(2)	7626(1)	5849(2)	35(1)
OW1	6338(2)	6342(1)	10399(2)	38(1)	N14	8009(3)	8224(1)	6598(3)	31(1)
OW2	14435(3)	8323(1)	5598(3)	54(1)	C14	8532(3)	7760(1)	5876(3)	30(1)
OW3	15755(3)	7868(1)	5777(3)	58(1)	C24	8518(3)	8099(1)	5777(3)	33(1)
OW4	14613(2)	7427(1)	5177(3)	45(1)	C34	9508(4)	8245(1)	5643(4)	41(1)
OW5	4356(2)	5164(1)	2598(3)	47(1)	C44	9475(5)	8569(1)	5363(5)	55(1)
OW6	14926(2)	6098(1)	1624(2)	36(1)	C54	8906(7)	8625(2)	4416(6)	79(2)
OW7	14003(3)	5576(1)	1088(3)	47(1)	C64	10510(6)	8688(2)	5332(6)	76(2)
OW8	11705(3)	7420(1)	9344(3)	53(1)	C15	10272(3)	7221(1)	6746(3)	29(1)
OW15	11550(3)	5390(1)	3002(3)	65(1)	O15	10737(2)	7410(1)	7201(2)	36(1)
OW23	14265(3)	4627(2)	6988(4)	94(2)	C25	9516(3)	7302(1)	5956(3)	30(1)
OW10	4141(5)	8592(2)	7404(5)	78(1)	C35	9835(3)	7200(1)	4971(3)	34(1)
OW11	5574(6)	9528(2)	6900(8)	78(1)	N15	9388(3)	7626(1)	5958(3)	30(1)
OW12	4365(5)	3764(2)	5300(9)	78(1)	C16	10562(3)	6586(1)	8290(3)	29(1)
OW13	3879(5)	4356(2)	5049(7)	78(1)	O16	11056(2)	6386(1)	8663(2)	40(1)
OW14	6438(9)	8476(3)	1858(9)	78(1)	N16	10368(3)	6933(1)	6904(3)	31(1)
OW16	6557(9)	8605(3)	2858(9)	78(1)	C26	11019(3)	6810(1)	7637(3)	32(1)
OW17	5487(12)	8897(4)	2704(12)	78(1)	C17	8272(3)	7010(1)	10897(3)	26(1)
OW18	3973(14)	4385(4)	4471(18)	78(1)	O17	8433(2)	7062(1)	11764(2)	30(1)
OW19	5969(17)	4285(5)	-853(17)	78(1)	C27	8920(3)	6803(1)	10337(3)	28(1)
OW21	5732(14)	9567(4)	7422(19)	78(1)	C47	8370(3)	6590(1)	9665(3)	27(1)
OW22	3820(20)	8660(7)	7100(20)	78(1)	O37	8247(2)	6304(1)	7441(2)	35(1)
OW24	7220(30)	4870(7)	-650(30)	78(1)	C57	9023(3)	6416(1)	8983(3)	27(1)
OW25	6310(20)	4773(7)	-290(20)	78(1)	C67	8341(3)	6240(1)	8304(3)	28(1)
OW28	6290(20)	8838(8)	2290(20)	78(1)	N17	9607(3)	6616(1)	8422(3)	28(1)
OW29	4500(30)	3857(9)	4740(30)	78(1)	O18	5985(3)	6110(1)	6834(2)	42(1)
OW30	6160(30)	8639(9)	2230(30)	78(1)	N18	7792(3)	6030(1)	8678(3)	29(1)
OW31	5470(20)	4806(7)	140(20)	78(1)	C18	6322(3)	6129(1)	7668(3)	32(1)
OW32	5520(30)	9579(9)	8240(30)	78(1)	O28	4876(4)	5416(2)	9385(3)	81(2)
OW33	6500(20)	4814(7)	1240(20)	78(1)	C28	7052(3)	5893(1)	8051(3)	32(1)
OW34	4280(30)	3699(9)	5790(40)	78(1)	O38	4176(3)	5357(1)	7959(3)	65(1)
OW35	6326(12)	4382(4)	-26(12)	78(1)	C38	6555(4)	5649(1)	8600(4)	42(1)
OW37	5250(40)	8803(11)	1950(40)	78(1)	C48	5792(4)	5488(2)	7962(4)	52(1)
OW38	12970(30)	4194(8)	6670(30)	78(1)	C58	4918(4)	5417(1)	8512(4)	47(1)
OW39	5860(40)	5056(12)	-600(40)	78(1)	O19	6140(2)	6948(1)	9025(2)	35(1)
OW40	6920(30)	4938(11)	-220(40)	78(1)	N19	6048(3)	6333(1)	8288(3)	29(1)
OW41	6700(30)	5088(11)	-740(40)	78(1)	C19	5942(3)	6865(1)	8200(3)	29(1)
C11	7376(3)	7666(1)	10538(3)	30(1)	C29	5388(3)	6574(1)	8030(3)	29(1)
O11	8216(2)	7721(1)	10753(3)	41(1)	C39	4466(3)	6569(1)	8631(4)	39(1)
C21	6923(3)	7367(1)	10777(3)	29(1)	C59	3940(4)	6269(1)	8587(4)	48(1)
N11	7529(3)	7132(1)	10401(3)	29(1)	C69	3545(4)	6172(1)	7625(5)	59(2)
N22	6782(3)	7860(1)	10071(3)	32(1)	C79	3803(4)	6828(1)	8345(4)	45(1)
C12	6881(3)	8193(1)	8692(3)	31(1)	O110	8207(2)	7043(1)	7832(2)	36(1)
O12	6429(2)	8004(1)	8217(2)	37(1)	N110	6200(3)	7017(1)	7423(3)	29(1)

Supplementary Table S10b. Atomic coordinates (x 10^4) and equivalent isotropic displacement parameters (Å^{2}x 10^3) for -4aa truncation of xanthomonin I. U(eq) is defined as one third of the trace of the orthogonalized U^{ij} tensor.

	x	y	z	U(eq)		x	y	z	U(eq)
C110	7740(3)	7271(1)	7913(3)	27(1)	O119	6302(3)	4690(1)	6885(3)	59(1)
C210	6698(3)	7297(1)	7528(3)	33(1)	N119	6682(3)	4514(1)	4344(3)	34(1)
O111	8941(2)	8059(1)	8304(2)	33(1)	C119	6741(4)	4662(1)	6110(4)	41(1)
N111	8093(2)	7517(1)	8323(3)	28(1)	C219	6281(3)	4475(1)	5316(4)	39(1)
C111	9472(3)	7850(1)	8556(3)	27(1)	C319	6464(5)	4142(1)	5523(5)	54(1)
C211	9070(3)	7542(1)	8724(3)	31(1)	C519	7152(4)	4244(1)	3971(4)	44(1)
O112	12300(3)	8381(1)	7922(3)	52(1)	C4A19	7277(8)	4068(2)	4911(9)	43(2)
N112	10427(3)	7884(1)	8729(3)	28(1)	C4B19	6661(14)	4000(3)	4560(11)	70(4)
C112	11718(3)	8174(1)	7910(4)	33(1)	O120	7645(2)	5391(1)	6488(2)	36(1)
C212	10911(3)	8171(1)	8652(3)	31(1)	N120	7623(3)	4766(1)	5970(3)	33(1)
C312	11287(4)	8280(1)	9642(4)	40(1)	C120	8365(3)	5231(1)	6527(3)	30(1)
C412	10486(3)	8407(1)	10246(3)	34(1)	C220	8242(3)	4897(1)	6717(3)	34(1)
C512	10290(4)	8708(1)	10215(4)	46(1)	C320	9189(4)	4722(1)	6828(3)	37(1)
C612	9532(5)	8829(1)	10749(5)	55(1)	C420	9091(5)	4396(1)	7085(5)	62(2)
C712	8989(4)	8647(2)	11313(4)	55(2)	C520	10093(6)	4253(2)	7081(5)	74(2)
C812	9175(4)	8346(1)	11344(4)	48(1)	C620	8620(6)	4350(2)	8048(9)	107(4)
C912	9914(4)	8226(1)	10811(3)	41(1)	O121	10645(2)	5516(1)	5143(2)	35(1)
O113	11300(2)	7774(1)	5381(2)	39(1)	N121	9264(3)	5333(1)	6413(3)	32(1)
N213	14610(3)	7774(1)	7582(4)	48(1)	C121	10162(3)	5712(1)	5530(3)	26(1)
N113	11730(3)	7951(1)	7278(3)	29(1)	C221	9484(3)	5650(1)	6358(3)	29(1)
C113	12094(3)	7890(1)	5568(3)	32(1)	C321	9950(4)	5758(1)	7306(3)	38(1)
C213	12502(3)	7915(1)	6595(3)	30(1)	O122	10722(2)	6570(1)	3654(2)	32(1)
O313	13405(3)	7627(1)	8528(3)	50(1)	N122	10207(3)	5998(1)	5284(3)	30(1)
C313	13094(3)	7629(1)	6828(4)	37(1)	C122	10278(3)	6351(1)	3945(3)	27(1)
C413	13725(3)	7674(1)	7715(4)	37(1)	C222	10809(3)	6113(1)	4533(3)	31(1)
O114	13013(2)	7442(1)	3967(2)	41(1)	O123	8050(2)	6628(1)	4735(2)	33(1)
N114	12670(3)	7984(1)	4870(3)	36(1)	N123	9340(3)	6304(1)	3764(2)	27(1)
C114	12397(3)	7612(1)	3608(3)	34(1)	C123	8125(3)	6690(1)	3881(3)	26(1)
C214	12368(4)	7940(1)	3870(3)	36(1)	O223	7959(2)	5812(1)	549(2)	33(1)
C314	12998(5)	8119(1)	3184(4)	49(1)	C223	8732(3)	6506(1)	3196(3)	26(1)
C414	12829(7)	8450(1)	3339(5)	75(2)	C323	8016(3)	6326(1)	2547(3)	28(1)
C514	14054(5)	8036(2)	3240(5)	66(2)	C423	8501(3)	6100(1)	1909(3)	29(1)
O115	12362(3)	7104(1)	975(2)	47(1)	C523	7836(3)	5880(1)	1393(3)	27(1)
N115	11774(3)	7517(1)	2918(3)	32(1)	O124	5852(2)	6932(1)	5351(2)	30(1)
C115	12557(4)	7156(1)	1823(3)	36(1)	O324	7391(4)	8083(1)	3095(5)	82(2)
C215	11774(3)	7202(1)	2577(3)	35(1)	N124	7587(3)	6911(1)	3490(2)	27(1)
C315	10755(4)	7162(1)	2167(4)	40(1)	C124	6186(3)	6870(1)	4555(3)	26(1)
C415	10490(4)	7467(1)	1785(4)	41(1)	O224	6139(4)	7845(1)	2435(3)	64(1)
C515	10918(4)	7680(1)	2540(4)	40(1)	C224	6948(3)	7073(1)	4118(3)	28(1)
O116	15647(3)	6814(1)	1783(3)	46(1)	C324	6506(3)	7336(1)	3582(3)	32(1)
N116	13463(3)	7177(1)	2169(3)	38(1)	C424	7287(4)	7572(1)	3383(4)	40(1)
C116	14787(3)	6851(1)	1624(3)	35(1)	C524	6867(5)	7846(1)	2935(4)	49(1)
O216	14158(2)	6634(1)	1474(3)	43(1)	O125	5907(2)	6030(1)	3340(2)	28(1)
C216	14320(4)	7154(1)	1577(4)	41(1)	N125	5898(2)	6632(1)	4053(2)	25(1)
O117	7841(2)	5162(1)	1712(3)	41(1)	C125	5827(3)	6100(1)	4205(3)	25(1)
N117	7141(3)	5756(1)	1918(3)	28(1)	C225	4270(3)	6388(1)	4068(3)	30(1)
C117	6999(3)	5219(1)	1876(3)	31(1)	C325	3736(3)	6109(1)	4410(4)	41(1)
C217	6542(3)	5513(1)	1565(3)	31(1)	C425	3726(3)	6674(1)	4304(4)	35(1)
O118	6073(2)	4973(1)	4214(2)	35(1)	C525	3458(4)	6715(1)	5359(4)	44(1)
N118	6408(3)	5035(1)	2347(3)	33(1)	C625	5326(3)	6391(1)	4461(3)	26(1)
C118	6468(3)	4758(1)	3839(3)	32(1)	O126	8041(2)	5871(1)	4367(3)	37(1)
C218	6748(3)	4762(1)	2788(3)	34(1)	N126	6139(2)	5930(1)	4916(2)	25(1)

Supplementary Table S10c. Atomic coordinates (x 10^4) and equivalent isotropic displacement parameters (Å^{2}x 10^3) for -4aa truncation of xanthomonin I. U(eq) is defined as one third of the trace of the orthogonalized U^{ij} tensor.

	x	y	z	U(eq)		x	y	z	U(eq)
C126	7568(3)	5642(1)	4373(3)	26(1)	C329	13556(3)	5259(1)	5044(4)	38(1)
C226	6524(3)	5635(1)	4724(3)	27(1)	C429	12318(3)	4997(1)	6013(3)	30(1)
O127	8685(2)	4840(1)	4287(2)	34(1)	O130	12573(3)	5416(1)	8829(3)	47(1)
N127	7885(3)	5383(1)	4065(3)	31(1)	N130	12444(3)	4889(1)	7735(3)	34(1)
C127	9225(3)	5030(1)	3976(3)	28(1)	C130	11944(3)	5225(1)	8992(3)	35(1)
C227	8849(3)	5335(1)	3691(3)	29(1)	C230	12100(4)	4905(1)	8728(3)	37(1)
O128	11954(3)	4484(1)	4910(3)	54(1)	C330	12833(4)	4751(1)	9429(4)	46(1)
N128	10175(3)	4986(1)	3839(3)	29(1)	C430	12919(6)	4423(2)	9179(5)	66(2)
C128	11452(3)	4709(1)	4770(3)	33(1)	C530	12552(5)	4794(2)	10473(4)	64(2)
C228	10631(3)	4699(1)	4024(3)	32(1)	O131	12028(2)	5615(1)	11053(2)	41(1)
C328	10973(4)	4560(1)	3072(4)	40(1)	N131	11135(3)	5308(1)	9440(3)	38(1)
C428	10134(4)	4457(1)	2454(3)	38(1)	C131	11689(3)	5756(1)	10350(3)	37(1)
C528	9707(5)	4638(1)	1749(4)	48(1)	C231	10962(4)	5621(1)	9651(4)	45(1)
C628	8897(5)	4545(2)	1206(4)	63(2)	C331	9961(4)	5620(2)	10085(5)	66(2)
C728	8509(5)	4278(2)	1359(5)	64(2)	C531	10291(4)	5123(2)	9631(5)	59(2)
C828	8911(5)	4089(1)	2052(5)	58(2)	C631	9422(8)	5313(3)	9739(9)	39(4)
C928	9701(4)	4182(1)	2604(4)	47(1)	C4A31	9855(7)	5355(2)	10428(8)	47(3)
O129	11085(2)	5103(1)	7170(2)	37(1)	O132	10759(3)	6272(1)	11612(3)	63(1)
N129	11566(3)	4959(1)	5284(3)	29(1)	N132	11877(3)	6042(1)	10193(3)	39(1)
C129	11882(3)	4999(1)	7027(3)	31(1)	C132	11613(3)	6316(1)	11681(4)	39(1)
O229	13317(3)	5315(1)	4207(3)	54(1)	O232	12059(2)	6456(1)	12407(2)	40(1)
N229	14430(3)	5160(1)	5296(4)	54(1)	C232	12339(4)	6230(1)	10921(3)	40(1)
C229	12861(3)	5291(1)	5860(3)	32(1)					

Supplementary Table S11a. Bond lengths [Å] and angles [°] for the -4 aa truncation of xanthomonin I.

O(100)-H(101)	0.9103	OH2-HH2	0.8400	C13-C23	1.516(8)
O(100)-H(102)	0.9813	C11-O11	1.215(6)	C23-C33	1.526(8)
OH1-C(100)	1.393(8)	C11-N22	1.353(6)	C23-H23	1.0000
OH1-H(6)	0.8400	C11-C21	1.527(6)	C33-C43	1.489(12)
C(100)-C(101)	1.559(11)	C21-N11	1.455(6)	C33-H3A3	0.9900
C(100)-H(10A)	0.9900	C21-H2A1	0.9900	C33-H3B3	0.9900
C(100)-H(10B)	0.9900	C21-H2B1	0.9900	C43-C53	1.505(9)
C(101)-C(102)	1.503(11)	N11-C17	1.340(6)	C43-H4A3	0.9900
C(101)-H(10C)	0.9900	N11-H11	0.8800	C43-H4B3	0.9900
C(101)-H(10D)	0.9900	N22-C22	1.442(6)	C53-H5A3	0.9900
C(102)-C(103)	1.546(13)	N22-H22	0.8800	C53-H5B3	0.9900
C(102)-H(10E)	0.9900	C12-O12	1.238(6)	O14-C14	1.242(6)
C(102)-H(10F)	0.9900	C12-N13	1.330(6)	N14-C24	1.468(6)
C(103)-C(104)	1.555(14)	C12-C22	1.514(7)	N14-H14	0.8800
C(103)-H(10G)	0.9900	C22-H2A2	0.9900	C14-N15	1.331(6)
C(103)-H(10H)	0.9900	C22-H2B2	0.9900	C14-C24	1.532(7)
C(104)-C(105)	1.552(14)	O13-C13	1.238(7)	C24-C34	1.532(7)
C(104)-H(10I)	0.9900	N13-C53	1.468(7)	C24-H24	1.0000
C(104)-H(10J)	0.9900	N13-C23	1.477(7)	C34-C44	1.514(8)
C(105)-OH2	1.465(14)	C13-N14	1.327(6)	C34-H3A4	0.9900

Supplementary Table S11b. Bond lengths [Å] and angles [°] for the -4 aa truncation of xanthomonin I.

C34-H3B4	0.9900	C48-C58	1.479(8)	C812-C912	1.387(8)
C44-C64	1.527(10)	C48-H4A8	0.9900	C812-H812	0.9500
C44-C54	1.535(11)	C48-H4B8	0.9900	C912-H912	0.9500
C44-H44	1.0000	O19-C19	1.230(5)	O113-C113	1.235(6)
C54-H5A4	0.9800	N19-C29	1.458(6)	N213-C413	1.319(7)
C54-H5B4	0.9800	N19-H19	0.8800	N213-H2A13	0.8800
C54-H5C4	0.9800	C19-N110	1.333(6)	N213-H2B13	0.8800
C64-H6A4	0.9800	C19-C29	1.534(6)	N113-C213	1.453(6)
C64-H6B4	0.9800	C29-C39	1.540(6)	N113-H113	0.8800
C64-H6C4	0.9800	C29-H29	1.0000	C113-N114	1.338(6)
C15-O15	1.230(5)	C39-C79	1.534(7)	C113-C213	1.524(6)
C15-N16	1.324(6)	C39-C59	1.535(8)	C213-C313	1.555(7)
C15-C25	1.540(6)	C39-H39	1.0000	C213-H213	1.0000
C25-N15	1.468(6)	C59-C69	1.495(9)	O313-C413	1.241(6)
C25-C35	1.520(6)	C59-H5A9	0.9900	C313-C413	1.504(7)
C25-H25	1.0000	C59-H5B9	0.9900	C313-H3A13	0.9900
C35-H3A5	0.9800	C69-H6A9	0.9800	C313-H3B13	0.9900
C35-H3B5	0.9800	C69-H6B9	0.9800	O114-C114	1.239(6)
C35-H3C5	0.9800	C69-H6C9	0.9800	N114-C214	1.455(6)
N15-H15	0.8800	C79-H7A9	0.9800	N114-H114	0.8800
C16-O16	1.234(6)	C79-H7B9	0.9800	C114-N115	1.342(6)
C16-N17	1.343(6)	C79-H7C9	0.9800	C114-C214	1.526(7)
C16-C26	1.506(7)	O110-C110	1.221(6)	C214-C314	1.535(7)
N16-C26	1.452(6)	N110-C210	1.444(6)	C214-H214	1.0000
N16-H16	0.8800	N110-H110	0.8800	C314-C514	1.504(10)
C26-H2A6	0.9900	C110-N111	1.335(6)	C314-C414	1.524(9)
C26-H2B6	0.9900	C110-C210	1.527(6)	C314-H314	1.0000
C17-O17	1.241(5)	C210-H2A10	0.9900	C414-H4A14	0.9800
C17-C27	1.520(6)	C210-H2B10	0.9900	C414-H4B14	0.9800
C27-C47	1.528(6)	O111-C111	1.237(5)	C414-H4C14	0.9800
C27-H2A7	0.9900	N111-C211	1.451(6)	C514-H5A14	0.9800
C27-H2B7	0.9900	N111-H111	0.8800	C514-H5B14	0.9800
C47-C57	1.540(6)	C111-N112	1.343(6)	C514-H5C14	0.9800
C47-H4A7	0.9900	C111-C211	1.516(6)	O115-C115	1.225(6)
C47-H4B7	0.9900	C211-H2A11	0.9900	N115-C515	1.478(6)
O37-C67	1.237(5)	C211-H2B11	0.9900	N115-C215	1.500(6)
C57-N17	1.449(6)	O112-C112	1.233(6)	C115-N116	1.333(7)
C57-C67	1.538(6)	N112-C212	1.462(6)	C115-C215	1.537(7)
C57-H57	1.0000	N112-H112	0.8800	C215-C315	1.515(7)
C67-N18	1.325(6)	C112-N113	1.336(6)	C215-H215	1.0000
N17-H17	0.8800	C112-C212	1.536(6)	C315-C415	1.514(7)
O18-C18	1.240(6)	C212-C312	1.538(7)	C315-H3A15	0.9900
N18-C28	1.463(6)	C212-H212	1.0000	C315-H3B15	0.9900
N18-H18	0.8800	C312-C412	1.516(7)	C415-C515	1.531(7)
C18-N19	1.322(6)	C312-H3A12	0.9900	C415-H4A15	0.9900
C18-C28	1.548(7)	C312-H3B12	0.9900	C415-H4B15	0.9900
O28-C58	1.215(7)	C412-C512	1.384(8)	C515-H5A15	0.9900
C28-C38	1.513(7)	C412-C912	1.392(7)	C515-H5B15	0.9900
C28-H2A8	1.0000	C512-C612	1.409(8)	O116-C116	1.212(6)
O38-C58	1.293(7)	C512-H512	0.9500	N116-C216	1.461(6)
O38-H38	0.8400	C612-C712	1.371(9)	N116-H116	0.8800
C38-C48	1.542(8)	C612-H612	0.9500	C116-O216	1.321(6)
C38-H3A8	0.9900	C712-C812	1.380(10)	C116-C216	1.513(7)
C38-H3B8	0.9900	C712-H712	0.9500	O216-H216	0.8400

Supplementary Table S11c. Bond lengths [Å] and angles [°] for the -4 aa truncation of xanthomonin I.

C216-H2A16	0.9900	C520-H5C20	0.9800	C125-N126	1.317(6)
C216-H2B16	0.9900	C620-H6A20	0.9800	C125-C625	1.527(6)
O117-C117	1.219(6)	C620-H6B20	0.9800	C225-C425	1.534(6)
N117-C523	1.342(6)	C620-H6C20	0.9800	C225-C325	1.537(7)
N117-C217	1.452(6)	O121-C121	1.241(6)	C225-C625	1.544(6)
N117-H117	0.8800	N121-C221	1.461(6)	C225-H225	1.0000
C117-N118	1.347(6)	N121-H121	0.8800	C325-H3A25	0.9800
C117-C217	1.523(6)	C121-N122	1.331(6)	C325-H3B25	0.9800
C217-H2A17	0.9900	C121-C221	1.527(6)	C325-H3C25	0.9800
C217-H2B17	0.9900	C221-C321	1.531(6)	C425-C525	1.530(7)
O118-C118	1.236(6)	C221-H221	1.0000	C425-H4A25	0.9900
N118-C218	1.449(6)	C321-H3A21	0.9800	C425-H4B25	0.9900
N118-H118	0.8800	C321-H3B21	0.9800	C525-H5A25	0.9800
C118-N119	1.334(6)	C321-H3C21	0.9800	C525-H5B25	0.9800
C118-C218	1.519(7)	O122-C122	1.237(5)	C525-H5C25	0.9800
C218-H2A18	0.9900	N122-C222	1.447(6)	C625-H625	1.0000
C218-H2B18	0.9900	N122-H122	0.8800	O126-C126	1.221(5)
O119-C119	1.254(6)	C122-N123	1.330(6)	N126-C226	1.458(6)
N119-C519	1.481(7)	C122-C222	1.525(6)	N126-H126	0.8800
N119-C219	1.483(6)	C222-H2A22	0.9900	C126-N127	1.325(6)
C119-N120	1.323(6)	C222-H2B22	0.9900	C126-C226	1.533(6)
C119-C219	1.516(8)	O123-C123	1.225(5)	C226-H2A26	0.9900
C219-C319	1.549(8)	N123-C223	1.457(5)	C226-H2B26	0.9900
C219-H219	1.0000	N123-H123	0.8800	O127-C127	1.224(6)
C319-C4A19	1.463(13)	C123-N124	1.349(6)	N127-C227	1.457(6)
C319-C4B19	1.515(17)	C123-C223	1.529(6)	N127-H127	0.8800
C319-H3A19	0.9900	O223-C523	1.228(5)	C127-N128	1.345(6)
C319-H3B19	0.9900	C223-C323	1.551(6)	C127-C227	1.516(6)
C319-H3A19	0.9900	C223-H223	1.0000	C227-H2A27	0.9900
C319-H3B19	0.9900	C323-C423	1.518(7)	C227-H2B27	0.9900
C519-C4A19	1.534(12)	C323-H3A23	0.9900	O128-C128	1.243(6)
C519-C4B19	1.540(15)	C323-H3B23	0.9900	N128-C228	1.460(6)
C519-H5A19	0.9900	C423-C523	1.517(6)	N128-H128	0.8800
C519-H5B19	0.9900	C423-H4A23	0.9900	C128-N129	1.340(6)
C519-H5A19	0.9900	C423-H4B23	0.9900	C128-C228	1.517(7)
C519-H5B19	0.9900	O124-C124	1.238(5)	C228-C328	1.548(6)
C4A19-H4AA19	0.9900	O324-C524	1.308(8)	C228-H228	1.0000
C4A19-H4AB19	0.9900	O324-H324	0.8400	C328-C428	1.498(8)
C4B19-H4BA19	0.9900	N124-C224	1.455(6)	C328-H3A28	0.9900
C4B19-H4BB19	0.9900	N124-H124	0.8800	C328-H3B28	0.9900
O120-C120	1.228(6)	C124-N125	1.339(6)	C428-C528	1.392(8)
N120-C220	1.453(6)	C124-C224	1.531(6)	C428-C928	1.398(8)
N120-H120	0.8800	O224-C524	1.207(8)	C528-C628	1.397(9)
C120-N121	1.336(6)	C224-C324	1.518(6)	C528-H528	0.9500
C120-C220	1.540(7)	C224-H224	1.0000	C628-C728	1.339(11)
C220-C320	1.532(7)	C324-C424	1.546(7)	C628-H628	0.9500
C220-H220	1.0000	C324-H3A24	0.9900	C728-C828	1.389(10)
C320-C420	1.519(7)	C324-H3B24	0.9900	C728-H728	0.9500
C320-H3A20	0.9900	C424-C524	1.493(8)	C828-C928	1.382(9)
C320-H3B20	0.9900	C424-H4A24	0.9900	C828-H828	0.9500
C420-C620	1.515(13)	C424-H4B24	0.9900	C928-H928	0.9500
C420-C520	1.526(10)	O125-C125	1.249(5)	O129-C129	1.216(6)
C420-H420	1.0000	N125-C625	1.464(6)	N129-C429	1.443(6)
C520-H5A20	0.9800	N125-H125	0.8800	N129-H129	0.8800

Supplementary Table S11d. Bond lengths [Å] and angles [°] for the -4 aa truncation of xanthomonin I.

C129-N130	1.333(6)	C132-C232	1.524(7)	O12-C12-C22	120.8(4)
C129-C429	1.545(6)	O232-H232	0.8400	N13-C12-C22	116.7(4)
O229-C329	1.226(7)	C232-H2A32	0.9900	N22-C22-C12	111.3(4)
N229-C329	1.326(7)	C232-H2B32	0.9900	N22-C22-H2A2	109.4
N229-H2A29	0.8800	H(101)-O(100)-H(102)	98.2	C12-C22-H2A2	109.4
N229-H2B29	0.8800	C(100)-OH1-H(6)	109.5	N22-C22-H2B2	109.4
C229-C329	1.509(7)	OH1-C(100)-C(101)	108.8(7)	C12-C22-H2B2	109.4
C229-C429	1.539(6)	OH1-C(100)-H(10A)	109.9	H2A2-C22-H2B2	108.0
C229-H2C29	0.9900	C(101)-C(100)-H(10A)	109.9	C12-N13-C53	125.2(4)
C229-H2D29	0.9900	OH1-C(100)-H(10B)	109.9	C12-N13-C23	121.9(4)
C429-H429	1.0000	C(101)-C(100)-H(10B)	109.9	C53-N13-C23	112.2(4)
O130-C130	1.247(6)	H(10A)-C(100)-H(10B)	108.3	O13-C13-N14	121.9(5)
N130-C230	1.472(6)	C(102)-C(101)-C(100)	114.0(9)	O13-C13-C23	118.4(5)
N130-H130	0.8800	C(102)-C(101)-H(10C)	108.7	N14-C13-C23	119.7(4)
C130-N131	1.345(6)	C(100)-C(101)-H(10C)	108.7	N13-C23-C13	117.7(4)
C130-C230	1.503(7)	C(102)-C(101)-H(10D)	108.7	N13-C23-C33	102.1(5)
C230-C330	1.551(7)	C(100)-C(101)-H(10D)	108.7	C13-C23-C33	110.7(5)
C230-H230	1.0000	H(10C)-C(101)-H(10D)	107.6	N13-C23-H23	108.7
C330-C430	1.522(9)	C(101)-C(102)-C(103)	110.4(12)	C13-C23-H23	108.7
C330-C530	1.523(8)	C(101)-C(102)-H(10E)	109.6	C33-C23-H23	108.7
C330-H330	1.0000	C(103)-C(102)-H(10E)	109.6	C43-C33-C23	104.5(5)
C430-H4A30	0.9800	C(101)-C(102)-H(10F)	109.6	C43-C33-H3A3	110.9
C430-H4B30	0.9800	C(103)-C(102)-H(10F)	109.6	C23-C33-H3A3	110.9
C430-H4C30	0.9800	H(10E)-C(102)-H(10F)	108.1	C43-C33-H3B3	110.9
C530-H5A30	0.9800	C(102)-C(103)-C(104)	130.0(17)	C23-C33-H3B3	110.9
C530-H5B30	0.9800	C(102)-C(103)-H(10G)	104.8	H3A3-C33-H3B3	108.9
C530-H5C30	0.9800	C(104)-C(103)-H(10G)	104.8	C33-C43-C53	104.7(6)
O131-C131	1.247(6)	C(102)-C(103)-H(10H)	104.8	C33-C43-H4A3	110.8
N131-C531	1.461(8)	C(104)-C(103)-H(10H)	104.8	C53-C43-H4A3	110.8
N131-C231	1.462(7)	H(10G)-C(103)-H(10H)	105.8	C33-C43-H4B3	110.8
C131-N132	1.332(7)	C(105)-C(104)-C(103)	123(2)	C53-C43-H4B3	110.8
C131-C231	1.510(7)	C(105)-C(104)-H(10I)	106.5	H4A3-C43-H4B3	108.9
C231-C331	1.521(8)	C(103)-C(104)-H(10I)	106.5	N13-C53-C43	102.9(5)
C231-H231	1.0000	C(105)-C(104)-H(10J)	106.5	N13-C53-H5A3	111.2
C331-C4A31	1.295(13)	C(103)-C(104)-H(10J)	106.5	C43-C53-H5A3	111.2
C331-C631	1.639(14)	H(10I)-C(104)-H(10J)	106.5	N13-C53-H5B3	111.2
C331-H3A31	0.9900	OH2-C(105)-C(104)	104(2)	C43-C53-H5B3	111.2
C331-H3B31	0.9900	C(105)-OH2-HH2	109.5	H5A3-C53-H5B3	109.1
C331-H3A31	0.9900	O11-C11-N22	123.4(4)	C13-N14-C24	122.0(4)
C331-H3B31	0.9900	O11-C11-C21	121.3(4)	C13-N14-H14	119.0
C531-C631	1.483(14)	N22-C11-C21	115.3(4)	C24-N14-H14	119.0
C531-C4A31	1.647(11)	N11-C21-C11	108.8(3)	O14-C14-N15	123.6(4)
C531-H5A31	0.9900	N11-C21-H2A1	109.9	O14-C14-C24	118.3(4)
C531-H5B31	0.9900	C11-C21-H2A1	109.9	N15-C14-C24	118.1(4)
C531-H5A31	0.9900	N11-C21-H2B1	109.9	N14-C24-C34	112.0(4)
C531-H5B31	0.9900	C11-C21-H2B1	109.9	N14-C24-C14	108.7(4)
C631-H6A31	0.9900	H2A1-C21-H2B1	108.3	C34-C24-C14	115.5(4)
C631-H6B31	0.9900	C17-N11-C21	123.7(4)	N14-C24-H24	106.7
C4A31-H4AA31	0.9900	C17-N11-H11	118.1	C34-C24-H24	106.7
C4A31-H4AB31	0.9900	C21-N11-H11	118.1	C14-C24-H24	106.7
O132-C132	1.197(6)	C11-N22-C22	121.7(4)	C44-C34-C24	115.2(5)
N132-C232	1.456(7)	C11-N22-H22	119.1	C44-C34-H3A4	108.5
N132-H132	0.8800	C22-N22-H22	119.1	C24-C34-H3A4	108.5
C132-O232	1.328(6)	O12-C12-N13	122.4(4)	C44-C34-H3B4	108.5

Supplementary Table S11e. Bond lengths [Å] and angles [°] for the -4 aa truncation of xanthomonin I.

C24-C34-H3B4	108.5	C17-C27-H2A7	108.7	O19-C19-N110	122.7(4)
H3A4-C34-H3B4	107.5	C47-C27-H2A7	108.7	O19-C19-C29	120.2(4)
C34-C44-C64	108.8(6)	C17-C27-H2B7	108.7	N110-C19-C29	117.1(4)
C34-C44-C54	113.0(5)	C47-C27-H2B7	108.7	N19-C29-C19	107.2(3)
C64-C44-C54	112.2(6)	H2A7-C27-H2B7	107.6	N19-C29-C39	111.9(4)
C34-C44-H44	107.5	C27-C47-C57	114.0(3)	C19-C29-C39	110.2(4)
C64-C44-H44	107.5	C27-C47-H4A7	108.7	N19-C29-H29	109.2
C54-C44-H44	107.5	C57-C47-H4A7	108.7	C19-C29-H29	109.2
C44-C54-H5A4	109.5	C27-C47-H4B7	108.7	C39-C29-H29	109.2
C44-C54-H5B4	109.5	C57-C47-H4B7	108.7	C79-C39-C59	112.6(4)
H5A4-C54-H5B4	109.5	H4A7-C47-H4B7	107.6	C79-C39-C29	110.0(4)
C44-C54-H5C4	109.5	N17-C57-C67	109.4(3)	C59-C39-C29	112.9(4)
H5A4-C54-H5C4	109.5	N17-C57-C47	110.8(4)	C79-C39-H39	107.0
H5B4-C54-H5C4	109.5	C67-C57-C47	106.4(3)	C59-C39-H39	107.0
C44-C64-H6A4	109.5	N17-C57-H57	110.1	C29-C39-H39	107.0
C44-C64-H6B4	109.5	C67-C57-H57	110.1	C69-C59-C39	117.0(5)
H6A4-C64-H6B4	109.5	C47-C57-H57	110.1	C69-C59-H5A9	108.0
C44-C64-H6C4	109.5	O37-C67-N18	120.1(4)	C39-C59-H5A9	108.0
H6A4-C64-H6C4	109.5	O37-C67-C57	121.3(4)	C69-C59-H5B9	108.0
H6B4-C64-H6C4	109.5	N18-C67-C57	118.4(4)	C39-C59-H5B9	108.0
O15-C15-N16	123.0(4)	C16-N17-C57	124.8(4)	H5A9-C59-H5B9	107.3
O15-C15-C25	122.5(4)	C16-N17-H17	117.6	C59-C69-H6A9	109.5
N16-C15-C25	114.5(4)	C57-N17-H17	117.6	C59-C69-H6B9	109.5
N15-C25-C35	110.0(4)	C67-N18-C28	117.7(4)	H6A9-C69-H6B9	109.5
N15-C25-C15	108.3(4)	C67-N18-H18	121.1	C59-C69-H6C9	109.5
C35-C25-C15	111.2(4)	C28-N18-H18	121.1	H6A9-C69-H6C9	109.5
N15-C25-H25	109.1	O18-C18-N19	123.3(4)	H6B9-C69-H6C9	109.5
C35-C25-H25	109.1	O18-C18-C28	119.9(4)	C39-C79-H7A9	109.5
C15-C25-H25	109.1	N19-C18-C28	116.6(4)	C39-C79-H7B9	109.5
C25-C35-H3A5	109.5	N18-C28-C38	109.0(4)	H7A9-C79-H7B9	109.5
C25-C35-H3B5	109.5	N18-C28-C18	110.8(4)	C39-C79-H7C9	109.5
H3A5-C35-H3B5	109.5	C38-C28-C18	111.9(4)	H7A9-C79-H7C9	109.5
C25-C35-H3C5	109.5	N18-C28-H2A8	108.4	H7B9-C79-H7C9	109.5
H3A5-C35-H3C5	109.5	C38-C28-H2A8	108.4	C19-N110-C210	120.2(4)
H3B5-C35-H3C5	109.5	C18-C28-H2A8	108.4	C19-N110-H110	119.9
C14-N15-C25	124.0(4)	C58-O38-H38	109.5	C210-N110-H110	119.9
C14-N15-H15	118.0	C28-C38-C48	111.4(4)	O110-C110-N111	123.5(4)
C25-N15-H15	118.0	C28-C38-H3A8	109.3	O110-C110-C210	121.9(4)
O16-C16-N17	123.4(4)	C48-C38-H3A8	109.3	N111-C110-C210	114.6(4)
O16-C16-C26	120.4(4)	C28-C38-H3B8	109.3	N110-C210-C110	114.2(4)
N17-C16-C26	116.2(4)	C48-C38-H3B8	109.3	N110-C210-H2A10	108.7
C15-N16-C26	123.4(4)	H3A8-C38-H3B8	108.0	C110-C210-H2A10	108.7
C15-N16-H16	118.3	C58-C48-C38	111.1(5)	N110-C210-H2B10	108.7
C26-N16-H16	118.3	C58-C48-H4A8	109.4	C110-C210-H2B10	108.7
N16-C26-C16	114.7(4)	C38-C48-H4A8	109.4	H2A10-C210-H2B10	107.6
N16-C26-H2A6	108.6	C58-C48-H4B8	109.4	C110-N111-C211	123.8(4)
C16-C26-H2A6	108.6	C38-C48-H4B8	109.4	C110-N111-H111	118.1
N16-C26-H2B6	108.6	H4A8-C48-H4B8	108.0	C211-N111-H111	118.1
C16-C26-H2B6	108.6	O28-C58-O38	122.4(5)	O111-C111-N112	122.6(4)
H2A6-C26-H2B6	107.6	O28-C58-C48	125.1(5)	O111-C111-C211	121.5(4)
O17-C17-N11	122.5(4)	O38-C58-C48	112.5(5)	N112-C111-C211	115.9(4)
O17-C17-C27	121.2(4)	C18-N19-C29	123.1(4)	N111-C211-C111	110.7(4)
N11-C17-C27	116.3(4)	C18-N19-H19	118.4	N111-C211-H2A11	109.5
C17-C27-C47	114.1(3)	C29-N19-H19	118.4	C111-C211-H2A11	109.5

Supplementary Table S11f. Bond lengths [Å] and angles [°] for the -4 aa truncation of xanthomonin I.

N111-C211-H2B11	109.5	C413-C313-C213	110.5(4)	C415-C315-H3B15	111.0
C111-C211-H2B11	109.5	C413-C313-H3A13	109.5	C215-C315-H3B15	111.0
H2A11-C211-H2B11	108.1	C213-C313-H3A13	109.5	H3A15-C315-H3B15	109.0
C111-N112-C212	122.3(4)	C413-C313-H3B13	109.5	C315-C415-C515	104.2(4)
C111-N112-H112	118.9	C213-C313-H3B13	109.5	C315-C415-H4A15	110.9
C212-N112-H112	118.9	H3A13-C313-H3B13	108.1	C515-C415-H4A15	110.9
O112-C112-N113	124.1(4)	O313-C413-N213	122.5(5)	C315-C415-H4B15	110.9
O112-C112-C212	118.7(4)	O313-C413-C313	120.7(4)	C515-C415-H4B15	110.9
N113-C112-C212	117.1(4)	N213-C413-C313	116.8(4)	H4A15-C415-H4B15	108.9
N112-C212-C112	113.3(4)	C113-N114-C214	119.2(4)	N115-C515-C415	102.8(4)
N112-C212-C312	111.2(4)	C113-N114-H114	120.4	N115-C515-H5A15	111.2
C112-C212-C312	111.3(4)	C214-N114-H114	120.4	C415-C515-H5A15	111.2
N112-C212-H212	106.9	O114-C114-N115	120.8(4)	N115-C515-H5B15	111.2
C112-C212-H212	106.9	O114-C114-C214	121.7(4)	C415-C515-H5B15	111.2
C312-C212-H212	106.9	N115-C114-C214	117.3(4)	H5A15-C515-H5B15	109.1
C412-C312-C212	112.3(4)	N114-C214-C114	110.6(4)	C115-N116-C216	123.9(4)
C412-C312-H3A12	109.1	N114-C214-C314	111.6(4)	C115-N116-H116	118.0
C212-C312-H3A12	109.1	C114-C214-C314	110.1(4)	C216-N116-H116	118.0
C412-C312-H3B12	109.1	N114-C214-H214	108.2	O116-C116-O216	124.3(5)
C212-C312-H3B12	109.1	C114-C214-H214	108.2	O116-C116-C216	123.1(4)
H3A12-C312-H3B12	107.9	C314-C214-H214	108.2	O216-C116-C216	112.6(4)
C512-C412-C912	118.7(5)	C514-C314-C414	112.8(6)	C116-O216-H216	109.5
C512-C412-C312	119.8(5)	C514-C314-C214	113.7(5)	N116-C216-C116	112.8(4)
C912-C412-C312	121.5(5)	C414-C314-C214	109.5(5)	N116-C216-H2A16	109.0
C412-C512-C612	120.8(5)	C514-C314-H314	106.8	C116-C216-H2A16	109.0
C412-C512-H512	119.6	C414-C314-H314	106.8	N116-C216-H2B16	109.0
C612-C512-H512	119.6	C214-C314-H314	106.8	C116-C216-H2B16	109.0
C712-C612-C512	119.5(6)	C314-C414-H4A14	109.5	H2A16-C216-H2B16	107.8
C712-C612-H612	120.3	C314-C414-H4B14	109.5	C523-N117-C217	122.7(4)
C512-C612-H612	120.3	H4A14-C414-H4B14	109.5	C523-N117-H117	118.6
C612-C712-C812	120.1(5)	C314-C414-H4C14	109.5	C217-N117-H117	118.6
C612-C712-H712	119.9	H4A14-C414-H4C14	109.5	O117-C117-N118	123.4(4)
C812-C712-H712	119.9	H4B14-C414-H4C14	109.5	O117-C117-C217	121.5(4)
C712-C812-C912	120.6(5)	C314-C514-H5A14	109.5	N118-C117-C217	115.0(4)
C712-C812-H812	119.7	C314-C514-H5B14	109.5	N117-C217-C117	109.4(3)
C912-C812-H812	119.7	H5A14-C514-H5B14	109.5	N117-C217-H2A17	109.8
C812-C912-C412	120.3(5)	C314-C514-H5C14	109.5	C117-C217-H2A17	109.8
C812-C912-H912	119.8	H5A14-C514-H5C14	109.5	N117-C217-H2B17	109.8
C412-C912-H912	119.8	H5B14-C514-H5C14	109.5	C117-C217-H2B17	109.8
C413-N213-H2A13	120.0	C114-N115-C515	126.0(4)	H2A17-C217-H2B17	108.2
C413-N213-H2B13	120.0	C114-N115-C215	121.6(4)	C117-N118-C218	122.4(4)
H2A13-N213-H2B13	120.0	C515-N115-C215	111.5(4)	C117-N118-H118	118.8
C112-N113-C213	122.4(4)	O115-C115-N116	122.8(5)	C218-N118-H118	118.8
C112-N113-H113	118.8	O115-C115-C215	122.7(4)	O118-C118-N119	121.3(4)
C213-N113-H113	118.8	N116-C115-C215	114.5(4)	O118-C118-C218	121.4(4)
O113-C113-N114	121.4(4)	N115-C215-C315	102.9(4)	N119-C118-C218	117.2(4)
O113-C113-C213	122.2(4)	N115-C215-C115	110.3(4)	N118-C218-C118	109.1(4)
N114-C113-C213	116.3(4)	C315-C215-C115	113.0(4)	N118-C218-H2A18	109.9
N113-C213-C113	111.0(4)	N115-C215-H215	110.2	C118-C218-H2A18	109.9
N113-C213-C313	110.3(4)	C315-C215-H215	110.2	N118-C218-H2B18	109.9
C113-C213-C313	108.4(4)	C115-C215-H215	110.2	C118-C218-H2B18	109.9
N113-C213-H213	109.0	C415-C315-C215	103.8(4)	H2A18-C218-H2B18	108.3
C113-C213-H213	109.0	C415-C315-H3A15	111.0	C118-N119-C519	125.9(4)
C313-C213-H213	109.0	C215-C315-H3A15	111.0	C118-N119-C219	119.7(4)

Supplementary Table S11g. Bond lengths [Å] and angles [°] for the -4 aa truncation of xanthomonin I.

C519-N119-C219	113.5(4)	C320-C220-C120	115.0(4)	C122-C222-H2A22	109.4
O119-C119-N120	123.9(5)	N120-C220-H220	107.0	N122-C222-H2B22	109.4
O119-C119-C219	118.4(4)	C320-C220-H220	107.0	C122-C222-H2B22	109.4
N120-C119-C219	117.4(4)	C120-C220-H220	107.0	H2A22-C222-H2B22	108.0
N119-C219-C119	115.9(4)	C420-C320-C220	116.2(5)	C122-N123-C223	123.2(4)
N119-C219-C319	102.8(4)	C420-C320-H3A20	108.2	C122-N123-H123	118.4
C119-C219-C319	110.0(4)	C220-C320-H3A20	108.2	C223-N123-H123	118.4
N119-C219-H219	109.3	C420-C320-H3B20	108.2	O123-C123-N124	120.0(4)
C119-C219-H219	109.3	C220-C320-H3B20	108.2	O123-C123-C223	122.6(4)
C319-C219-H219	109.3	H3A20-C320-H3B20	107.4	N124-C123-C223	117.0(4)
C4A19-C319-C219	103.7(5)	C620-C420-C320	112.5(6)	N123-C223-C123	108.6(3)
C4B19-C319-C219	106.2(6)	C620-C420-C520	110.8(5)	N123-C223-C323	109.8(3)
C4A19-C319-H3A19	111.0	C320-C420-C520	108.8(6)	C123-C223-C323	107.1(3)
C219-C319-H3A19	111.0	C620-C420-H420	108.2	N123-C223-H223	110.4
C4A19-C319-H3B19	111.0	C320-C420-H420	108.2	C123-C223-H223	110.4
C219-C319-H3B19	111.0	C520-C420-H420	108.2	C323-C223-H223	110.4
H3A19-C319-H3B19	109.0	C420-C520-H5A20	109.5	C423-C323-C223	114.0(4)
C4B19-C319-H3A19	110.5	C420-C520-H5B20	109.5	C423-C323-H3A23	108.8
C219-C319-H3A19	110.5	H5A20-C520-H5B20	109.5	C223-C323-H3A23	108.8
C4B19-C319-H3B19	110.5	C420-C520-H5C20	109.5	C423-C323-H3B23	108.8
C219-C319-H3B19	110.5	H5A20-C520-H5C20	109.5	C223-C323-H3B23	108.8
H3A19-C319-H3B19	108.7	H5B20-C520-H5C20	109.5	H3A23-C323-H3B23	107.7
N119-C519-C4A19	99.7(6)	C420-C620-H6A20	109.5	C523-C423-C323	116.3(4)
N119-C519-C4B19	101.5(7)	C420-C620-H6B20	109.5	C523-C423-H4A23	108.2
N119-C519-H5A19	111.8	H6A20-C620-H6B20	109.5	C323-C423-H4A23	108.2
C4A19-C519-H5A19	111.8	C420-C620-H6C20	109.5	C523-C423-H4B23	108.2
N119-C519-H5B19	111.8	H6A20-C620-H6C20	109.5	C323-C423-H4B23	108.2
C4A19-C519-H5B19	111.8	H6B20-C620-H6C20	109.5	H4A23-C423-H4B23	107.4
H5A19-C519-H5B19	109.6	C120-N121-C221	122.5(4)	O223-C523-N117	122.2(4)
N119-C519-H5A19	111.5	C120-N121-H121	118.8	O223-C523-C423	121.1(4)
C4B19-C519-H5A19	111.5	C221-N121-H121	118.8	N117-C523-C423	116.6(4)
N119-C519-H5B19	111.5	O121-C121-N122	123.4(4)	C524-O324-H324	109.5
C4B19-C519-H5B19	111.5	O121-C121-C221	122.6(4)	C123-N124-C224	117.8(3)
H5A19-C519-H5B19	109.3	N122-C121-C221	113.9(4)	C123-N124-H124	121.1
C319-C4A19-C519	107.8(7)	N121-C221-C121	110.7(3)	C224-N124-H124	121.1
C319-C4A19-H4AA19	110.1	N121-C221-C321	110.5(4)	O124-C124-N125	122.2(4)
C519-C4A19-H4AA19	110.1	C121-C221-C321	109.7(4)	O124-C124-C224	119.7(4)
C319-C4A19-H4AB19	110.1	N121-C221-H221	108.7	N125-C124-C224	118.2(3)
C519-C4A19-H4AB19	110.1	C121-C221-H221	108.7	N124-C224-C324	109.9(3)
H4AA19-C4A19-H4AB19	108.5	C321-C221-H221	108.7	N124-C224-C124	111.8(3)
C319-C4B19-C519	104.9(9)	C221-C321-H3A21	109.5	C324-C224-C124	112.8(4)
C319-C4B19-H4BA19	110.8	C221-C321-H3B21	109.5	N124-C224-H224	107.3
C519-C4B19-H4BA19	110.8	H3A21-C321-H3B21	109.5	C324-C224-H224	107.3
C319-C4B19-H4BB19	110.8	C221-C321-H3C21	109.5	C124-C224-H224	107.3
C519-C4B19-H4BB19	110.8	H3A21-C321-H3C21	109.5	C224-C324-C424	110.6(4)
H4BA19-C4B19-H4BB19	108.8	H3B21-C321-H3C21	109.5	C224-C324-H3A24	109.5
C119-N120-C220	124.4(4)	C121-N122-C222	124.3(4)	C424-C324-H3A24	109.5
C119-N120-H120	117.8	C121-N122-H122	117.9	C224-C324-H3B24	109.5
C220-N120-H120	117.8	C222-N122-H122	117.9	C424-C324-H3B24	109.5
O120-C120-N121	123.2(4)	O122-C122-N123	123.6(4)	H3A24-C324-H3B24	108.1
O120-C120-C220	119.2(4)	O122-C122-C222	120.1(4)	C524-C424-C324	112.3(5)
N121-C120-C220	117.7(4)	N123-C122-C222	116.2(4)	C524-C424-H4A24	109.1
N120-C220-C320	110.3(4)	N122-C222-C122	111.2(3)	C324-C424-H4A24	109.1
N120-C220-C120	110.0(4)	N122-C222-H2A22	109.4	C524-C424-H4B24	109.1

Supplementary Table S11h. Bond lengths [Å] and angles [°] for the -4 aa truncation of xanthomonin I.

C324-C424-H4B24	109.1	C126-N127-H127	118.0	C329-N229-H2B29	120.0
H4A24-C424-H4B24	107.9	C227-N127-H127	118.0	H2A29-N229-H2B29	120.0
O224-C524-O324	123.3(6)	O127-C127-N128	123.4(4)	C329-C229-C429	110.0(4)
O224-C524-C424	123.2(5)	O127-C127-C227	121.5(4)	C329-C229-H2C29	109.7
O324-C524-C424	113.4(6)	N128-C127-C227	115.1(4)	C429-C229-H2C29	109.7
C124-N125-C625	123.5(3)	N127-C227-C127	110.5(4)	C329-C229-H2D29	109.7
C124-N125-H125	118.3	N127-C227-H2A27	109.6	C429-C229-H2D29	109.7
C625-N125-H125	118.3	C127-C227-H2A27	109.6	H2C29-C229-H2D29	108.2
O125-C125-N126	122.6(4)	N127-C227-H2B27	109.6	O229-C329-N229	122.7(5)
O125-C125-C625	119.5(4)	C127-C227-H2B27	109.6	O229-C329-C229	122.1(5)
N126-C125-C625	117.9(4)	H2A27-C227-H2B27	108.1	N229-C329-C229	115.1(5)
C425-C225-C325	112.4(4)	C127-N128-C228	121.6(4)	N129-C429-C229	110.5(4)
C425-C225-C625	112.4(4)	C127-N128-H128	119.2	N129-C429-C129	110.5(4)
C325-C225-C625	110.6(4)	C228-N128-H128	119.2	C229-C429-C129	108.9(4)
C425-C225-H225	107.0	O128-C128-N129	123.2(4)	N129-C429-H429	108.9
C325-C225-H225	107.0	O128-C128-C228	119.0(4)	C229-C429-H429	108.9
C625-C225-H225	107.0	N129-C128-C228	117.7(4)	C129-C429-H429	108.9
C225-C325-H3A25	109.5	N128-C228-C128	113.9(4)	C129-N130-C230	118.6(4)
C225-C325-H3B25	109.5	N128-C228-C328	110.4(4)	C129-N130-H130	120.7
H3A25-C325-H3B25	109.5	C128-C228-C328	111.0(4)	C230-N130-H130	120.7
C225-C325-H3C25	109.5	N128-C228-H228	107.1	O130-C130-N131	118.8(5)
H3A25-C325-H3C25	109.5	C128-C228-H228	107.1	O130-C130-C230	120.8(4)
H3B25-C325-H3C25	109.5	C328-C228-H228	107.1	N131-C130-C230	120.3(4)
C525-C425-C225	115.8(4)	C428-C328-C228	111.5(4)	N130-C230-C130	109.1(4)
C525-C425-H4A25	108.3	C428-C328-H3A28	109.3	N130-C230-C330	110.2(4)
C225-C425-H4A25	108.3	C228-C328-H3A28	109.3	C130-C230-C330	111.9(4)
C525-C425-H4B25	108.3	C428-C328-H3B28	109.3	N130-C230-H230	108.5
C225-C425-H4B25	108.3	C228-C328-H3B28	109.3	C130-C230-H230	108.5
H4A25-C425-H4B25	107.4	H3A28-C328-H3B28	108.0	C330-C230-H230	108.5
C425-C525-H5A25	109.5	C528-C428-C928	116.7(5)	C430-C330-C530	111.4(5)
C425-C525-H5B25	109.5	C528-C428-C328	122.0(5)	C430-C330-C230	110.3(5)
H5A25-C525-H5B25	109.5	C928-C428-C328	121.1(5)	C530-C330-C230	111.3(5)
C425-C525-H5C25	109.5	C428-C528-C628	121.6(6)	C430-C330-H330	107.9
H5A25-C525-H5C25	109.5	C428-C528-H528	119.2	C530-C330-H330	107.9
H5B25-C525-H5C25	109.5	C628-C528-H528	119.2	C230-C330-H330	107.9
N125-C625-C125	107.1(3)	C728-C628-C528	120.1(6)	C330-C430-H4A30	109.5
N125-C625-C225	112.5(3)	C728-C628-H628	119.9	C330-C430-H4B30	109.5
C125-C625-C225	109.9(3)	C528-C628-H628	119.9	H4A30-C430-H4B30	109.5
N125-C625-H625	109.1	C628-C728-C828	120.5(6)	C330-C430-H4C30	109.5
C125-C625-H625	109.1	C628-C728-H728	119.8	H4A30-C430-H4C30	109.5
C225-C625-H625	109.1	C828-C728-H728	119.8	H4B30-C430-H4C30	109.5
C125-N126-C226	120.6(4)	C928-C828-C728	119.7(6)	C330-C530-H5A30	109.5
C125-N126-H126	119.7	C928-C828-H828	120.1	C330-C530-H5B30	109.5
C226-N126-H126	119.7	C728-C828-H828	120.1	H5A30-C530-H5B30	109.5
O126-C126-N127	124.3(4)	C828-C928-C428	121.3(6)	C330-C530-H5C30	109.5
O126-C126-C226	121.8(4)	C828-C928-H928	119.3	H5A30-C530-H5C30	109.5
N127-C126-C226	113.9(4)	C428-C928-H928	119.3	H5B30-C530-H5C30	109.5
N126-C226-C126	112.9(3)	C128-N129-C429	123.2(4)	C130-N131-C531	127.0(5)
N126-C226-H2A26	109.0	C128-N129-H129	118.4	C130-N131-C231	120.1(4)
C126-C226-H2A26	109.0	C429-N129-H129	118.4	C531-N131-C231	112.2(4)
N126-C226-H2B26	109.0	O129-C129-N130	122.3(4)	O131-C131-N132	123.5(5)
C126-C226-H2B26	109.0	O129-C129-C429	121.6(4)	O131-C131-C231	121.8(5)
H2A26-C226-H2B26	107.8	N130-C129-C429	116.0(4)	N132-C131-C231	114.6(4)
C126-N127-C227	124.0(4)	C329-N229-H2A29	120.0	N131-C231-C131	114.3(4)

Supplementary Table S11i. Bond lengths [Å] and angles [°] for the -4 aa truncation of xanthomonin I.

N131-C231-C331	103.4(5)	N131-C531-C4A31	93.8(6)	C331-C4A31-H4AA31	110.3
C131-C231-C331	110.0(4)	N131-C531-H5A31	109.7	C531-C4A31-H4AA31	110.3
N131-C231-H231	109.7	C631-C531-H5A31	109.7	C331-C4A31-H4AB31	110.3
C131-C231-H231	109.7	N131-C531-H5B31	109.7	C531-C4A31-H4AB31	110.3
C331-C231-H231	109.7	C631-C531-H5B31	109.7	H4AA31-C4A31-H4AB31	108.6
C4A31-C331-C231	105.1(7)	H5A31-C531-H5B31	108.2	C131-N132-C232	122.2(4)
C231-C331-C631	107.2(6)	N131-C531-H5A31	113.0	C131-N132-H132	118.9
C231-C331-H3A31	110.3	C4A31-C531-H5A31	113.0	C232-N132-H132	118.9
C631-C331-H3A31	110.3	N131-C531-H5B31	113.0	O132-C132-O232	125.3(5)
C231-C331-H3B31	110.3	C4A31-C531-H5B31	113.0	O132-C132-C232	124.4(5)
C631-C331-H3B31	110.3	H5A31-C531-H5B31	110.4	O232-C132-C232	110.3(4)
H3A31-C331-H3B31	108.5	C531-C631-C331	98.8(7)	C132-O232-H232	109.5
C4A31-C331-H3A31	110.7	C531-C631-H6A31	112.0	N132-C232-C132	110.2(4)
C231-C331-H3A31	110.7	C331-C631-H6A31	112.0	N132-C232-H2A32	109.6
C4A31-C331-H3B31	110.7	C531-C631-H6B31	112.0	C132-C232-H2A32	109.6
C231-C331-H3B31	110.7	C331-C631-H6B31	112.0	N132-C232-H2B32	109.6
H3A31-C331-H3B31	108.8	H6A31-C631-H6B31	109.7	C132-C232-H2B32	109.6
N131-C531-C631	110.0(7)	C331-C4A31-C531	107.0(7)	H2A32-C232-H2B32	108.1

Supplementary Table S12a. Anisotropic displacement parameters ($Å^2 x\ 10^3$) for the -4 aa truncation of xanthomonin I. The anisotropic displacement factor exponent takes the form: $-2\pi^2[\ h^2\ a^{*2}U^{11} + ...\ + 2\ h\ k\ a^*\ b^*\ U^{12}\ ]$.

	U11	U22	U33	U23		U11	U22	U33	U23
O(100)	36(2)	34(2)	25(1)	1(1)	C22	33(2)	33(2)	33(2)	2(2)
OH1	38(2)	100(3)	49(2)	10(2)	O13	82(3)	80(3)	44(2)	10(2)
C(100)	80(5)	103(6)	70(5)	-16(4)	N13	34(2)	29(2)	39(2)	1(2)
C(101)	129(8)	126(7)	114(7)	-25(6)	C13	38(3)	38(2)	39(3)	11(2)
C(102)	155(11)	113(7)	131(9)	-31(6)	C23	31(2)	44(3)	47(3)	13(2)
C(103)	190(16)	109(9)	310(20)	-2(11)	C33	91(5)	39(3)	66(4)	16(3)
C(104)	350(30)	340(30)	330(20)	160(20)	C43	110(6)	40(3)	75(5)	9(3)
C(105)	430(30)	370(30)	233(16)	158(18)	C53	51(3)	33(2)	50(3)	-4(2)
OH2	108(6)	87(5)	257(11)	28(6)	O14	27(2)	39(2)	38(2)	3(2)
OW1	37(2)	46(2)	31(2)	0(2)	N14	31(2)	35(2)	29(2)	4(2)
OW2	43(2)	53(2)	67(2)	-4(2)	C14	33(2)	38(2)	18(2)	0(2)
OW3	53(2)	50(2)	71(3)	-2(2)	C24	38(3)	33(2)	27(2)	5(2)
OW4	34(2)	43(2)	58(2)	-4(2)	C34	43(3)	40(3)	39(2)	2(2)
OW5	33(2)	65(2)	43(2)	11(2)	C44	64(4)	43(3)	59(3)	0(3)
OW6	33(2)	42(2)	35(2)	2(2)	C54	100(6)	63(4)	76(5)	29(4)
OW7	42(2)	52(2)	47(2)	8(2)	C64	87(5)	52(4)	91(5)	-14(4)
OW8	41(2)	74(3)	45(2)	20(2)	C15	28(2)	35(2)	23(2)	0(2)
OW15	52(2)	98(3)	45(2)	17(2)	O15	43(2)	38(2)	28(2)	1(1)
OW23	41(2)	164(6)	78(3)	4(4)	C25	29(2)	35(2)	26(2)	1(2)
C11	31(2)	33(2)	25(2)	-5(2)	C35	37(2)	39(2)	27(2)	-1(2)
O11	31(2)	35(2)	58(2)	5(2)	N15	28(2)	33(2)	28(2)	1(2)
C21	30(2)	33(2)	25(2)	2(2)	C16	26(2)	35(2)	27(2)	-4(2)
N11	32(2)	33(2)	23(2)	-2(2)	O16	32(2)	45(2)	42(2)	9(2)
N22	32(2)	38(2)	28(2)	4(2)	N16	29(2)	35(2)	27(2)	0(2)
C12	22(2)	33(2)	39(2)	1(2)	C26	23(2)	43(3)	30(2)	-1(2)
O12	35(2)	40(2)	36(2)	4(2)	C17	26(2)	28(2)	23(2)	0(2)

Supplementary Table S12b. Anisotropic displacement parameters ($Å^2 x\ 10^3$) for the -4 aa truncation of xanthomonin I. The anisotropic displacement factor exponent takes the form: $-2\pi^2[\ h^2\ a^{*2}U^{11} + ...\ + 2\ h\ k\ a^*\ b^*\ U^{12}\ ]$.

	U11	U22	U33	U23		U11	U22	U33	U23
O17	33(2)	37(2)	20(1)	2(1)	C114	32(2)	39(2)	30(2)	-4(2)
C27	23(2)	32(2)	27(2)	1(2)	C214	36(2)	38(2)	33(2)	-6(2)
C47	27(2)	31(2)	23(2)	-1(2)	C314	66(4)	42(3)	39(3)	-5(2)
O37	38(2)	45(2)	21(1)	1(1)	C414	123(6)	43(3)	59(4)	-2(3)
C57	27(2)	32(2)	23(2)	1(2)	C514	66(4)	84(5)	50(3)	-4(3)
C67	29(2)	28(2)	27(2)	0(2)	O115	52(2)	57(2)	32(2)	2(2)
N17	26(2)	30(2)	26(2)	2(2)	N115	33(2)	30(2)	33(2)	1(2)
O18	42(2)	61(2)	23(2)	-8(2)	C115	42(3)	34(2)	32(2)	4(2)
N18	31(2)	34(2)	23(2)	-2(2)	C215	39(3)	28(2)	38(2)	2(2)
C18	29(2)	41(2)	25(2)	-2(2)	C315	35(3)	36(2)	50(3)	3(2)
O28	59(3)	141(5)	43(2)	-1(3)	C415	37(3)	38(3)	47(3)	-1(2)
C28	28(2)	42(2)	24(2)	-7(2)	C515	37(3)	40(3)	42(3)	3(2)
O38	45(2)	101(4)	48(2)	-3(2)	O116	33(2)	44(2)	60(2)	-1(2)
C38	42(3)	47(3)	36(2)	-2(2)	N116	33(2)	45(2)	35(2)	3(2)
C48	51(3)	67(4)	39(3)	2(3)	C116	34(3)	38(2)	32(2)	1(2)
C58	46(3)	48(3)	46(3)	-3(3)	O216	34(2)	40(2)	55(2)	0(2)
O19	31(2)	51(2)	23(2)	-2(1)	C216	36(3)	43(3)	45(3)	6(2)
N19	27(2)	37(2)	23(2)	-3(2)	O117	32(2)	41(2)	49(2)	4(2)
C19	18(2)	39(2)	30(2)	-1(2)	N117	31(2)	28(2)	27(2)	-2(2)
C29	24(2)	39(2)	24(2)	2(2)	C117	34(2)	31(2)	29(2)	-2(2)
C39	30(2)	47(3)	40(3)	5(2)	C217	27(2)	36(2)	30(2)	3(2)
C59	33(3)	51(3)	60(3)	7(3)	O118	33(2)	35(2)	36(2)	0(2)
C69	42(3)	47(3)	87(5)	-6(3)	N118	27(2)	36(2)	35(2)	3(2)
C79	26(2)	51(3)	58(3)	-5(3)	C118	25(2)	31(2)	39(2)	1(2)
O110	29(2)	35(2)	44(2)	1(2)	C218	34(2)	29(2)	38(2)	-2(2)
N110	28(2)	38(2)	22(2)	1(2)	O119	44(2)	95(3)	40(2)	1(2)
C110	26(2)	35(2)	20(2)	9(2)	N119	34(2)	28(2)	40(2)	6(2)
C210	27(2)	35(2)	35(2)	5(2)	C119	34(3)	52(3)	36(2)	10(2)
O111	27(2)	39(2)	33(2)	2(1)	C219	30(2)	43(3)	43(3)	15(2)
N111	20(2)	34(2)	30(2)	2(2)	C319	62(4)	43(3)	57(3)	17(3)
C111	23(2)	38(2)	21(2)	2(2)	C519	46(3)	31(2)	56(3)	4(2)
C211	27(2)	37(2)	29(2)	5(2)	C4A19	38(6)	33(5)	58(7)	11(5)
O112	54(2)	39(2)	63(2)	-16(2)	C4B19	102(12)	38(6)	68(9)	14(6)
N112	26(2)	28(2)	28(2)	3(2)	O120	31(2)	40(2)	38(2)	-5(2)
C112	29(2)	28(2)	42(3)	0(2)	N120	28(2)	42(2)	29(2)	6(2)
C212	28(2)	32(2)	32(2)	-4(2)	C120	32(2)	38(2)	21(2)	3(2)
C312	34(3)	52(3)	35(2)	-9(2)	C220	36(2)	39(3)	28(2)	4(2)
C412	34(2)	42(3)	28(2)	-6(2)	C320	43(3)	37(2)	31(2)	2(2)
C512	55(3)	41(3)	41(3)	-7(2)	C420	77(4)	36(3)	72(4)	11(3)
C612	60(4)	47(3)	58(3)	-17(3)	C520	104(6)	55(4)	63(4)	7(3)
C712	43(3)	75(4)	48(3)	-29(3)	C620	70(5)	82(5)	170(10)	84(7)
C812	45(3)	65(4)	34(3)	-7(3)	O121	38(2)	34(2)	33(2)	-1(1)
C912	49(3)	44(3)	30(2)	-4(2)	N121	29(2)	32(2)	36(2)	3(2)
O113	29(2)	52(2)	36(2)	-3(2)	C121	25(2)	29(2)	25(2)	0(2)
N213	26(2)	64(3)	55(3)	4(2)	C221	29(2)	31(2)	28(2)	2(2)
N113	22(2)	36(2)	28(2)	1(2)	C321	44(3)	42(3)	27(2)	0(2)
C113	26(2)	35(2)	35(2)	-4(2)	O122	28(2)	30(2)	38(2)	1(1)
C213	27(2)	31(2)	32(2)	-5(2)	N122	28(2)	32(2)	29(2)	0(2)
O313	36(2)	72(3)	42(2)	7(2)	C122	24(2)	33(2)	23(2)	-5(2)
C313	33(2)	38(2)	41(3)	-8(2)	C222	23(2)	37(2)	33(2)	5(2)
C413	29(2)	36(2)	46(3)	0(2)	O123	31(2)	43(2)	24(2)	5(1)
O114	38(2)	43(2)	41(2)	-6(2)	N123	24(2)	31(2)	27(2)	9(2)
N114	32(2)	43(2)	32(2)	-8(2)	C123	21(2)	30(2)	26(2)	3(2)

Supplementary Table S12c. Anisotropic displacement parameters ($Å^2 x\ 10^3$) for the -4 aa truncation of xanthomonin I. The anisotropic displacement factor exponent takes the form: $-2\pi^2[\ h^2\ a^{*2}U^{11} + ...\ + 2\ h\ k\ a^*\ b^*\ U^{12}\ ]$.

	U11	U22	U33	U23		U11	U22	U33	U23
O223	37(2)	37(2)	25(2)	1(1)	C428	43(3)	36(2)	34(2)	-9(2)
C223	26(2)	30(2)	23(2)	3(2)	C528	70(4)	40(3)	35(3)	-7(2)
C323	26(2)	30(2)	28(2)	4(2)	C628	84(5)	60(4)	44(3)	-18(3)
C423	25(2)	34(2)	30(2)	1(2)	C728	60(4)	68(4)	62(4)	-34(4)
C523	25(2)	31(2)	27(2)	4(2)	C828	62(4)	50(3)	62(4)	-21(3)
O124	29(2)	38(2)	23(1)	-3(1)	C928	59(3)	43(3)	40(3)	-9(2)
O324	99(4)	34(2)	115(4)	5(2)	O129	26(2)	49(2)	34(2)	0(2)
N124	28(2)	31(2)	22(2)	-1(2)	N129	27(2)	28(2)	31(2)	-1(2)
C124	21(2)	32(2)	24(2)	0(2)	C129	29(2)	31(2)	31(2)	0(2)
O224	83(3)	52(2)	56(2)	15(2)	O229	55(2)	70(3)	38(2)	10(2)
C224	30(2)	30(2)	23(2)	-4(2)	N229	38(2)	63(3)	64(3)	7(3)
C324	38(2)	30(2)	30(2)	0(2)	C229	28(2)	34(2)	35(2)	1(2)
C424	55(3)	32(2)	32(2)	0(2)	C329	33(3)	35(2)	47(3)	-1(2)
C524	63(4)	39(3)	46(3)	4(3)	C429	27(2)	34(2)	30(2)	-1(2)
O125	25(2)	36(2)	24(2)	1(1)	O130	33(2)	53(2)	55(2)	-6(2)
N125	23(2)	30(2)	22(2)	1(2)	N130	25(2)	44(2)	33(2)	2(2)
C125	20(2)	30(2)	24(2)	3(2)	C130	28(2)	46(3)	32(2)	1(2)
C225	24(2)	33(2)	32(2)	-1(2)	C230	35(2)	46(3)	31(2)	5(2)
C325	26(2)	39(3)	59(3)	-5(2)	C330	41(3)	57(3)	39(3)	9(3)
C425	25(2)	36(2)	45(3)	-1(2)	C430	77(4)	69(4)	53(3)	15(3)
C525	36(3)	44(3)	54(3)	-5(3)	C530	76(4)	78(4)	38(3)	12(3)
C625	26(2)	29(2)	23(2)	2(2)	O131	36(2)	57(2)	31(2)	5(2)
O126	28(2)	31(2)	53(2)	1(2)	N131	35(2)	49(2)	31(2)	0(2)
N126	25(2)	30(2)	21(2)	-1(2)	C131	34(2)	51(3)	24(2)	-1(2)
C126	23(2)	32(2)	21(2)	4(2)	C231	44(3)	53(3)	37(3)	-9(2)
C226	25(2)	28(2)	27(2)	1(2)	C331	26(3)	98(5)	73(4)	-39(4)
O127	28(2)	38(2)	37(2)	1(2)	C531	48(3)	67(4)	62(4)	15(3)
N127	22(2)	33(2)	37(2)	-2(2)	C631	24(6)	58(8)	35(7)	-7(6)
C127	26(2)	35(2)	23(2)	-6(2)	C4A31	37(5)	52(6)	51(7)	-5(5)
C227	24(2)	39(2)	26(2)	1(2)	O132	31(2)	98(3)	60(2)	-31(2)
O128	51(2)	40(2)	70(3)	-12(2)	N132	37(2)	50(2)	29(2)	3(2)
N128	25(2)	32(2)	31(2)	-3(2)	C132	32(3)	46(3)	39(3)	-3(2)
C128	29(2)	32(2)	39(2)	-2(2)	O232	33(2)	51(2)	34(2)	-6(2)
C228	30(2)	31(2)	37(2)	-2(2)	C232	33(2)	57(3)	32(2)	-1(2)
C328	38(3)	43(3)	38(3)	-8(2)					

Supplementary Table S13a. Hydrogen coordinates (Å x 10^4) and isotropic displacement parameters (Å^2 x 10^3) for the -4 aa truncation of xanthomonin I.

	x	y	z	U(eq)		x	y	z	U(eq)
H(101)	8905	6433	6531	47	H3A8	6237	5735	9167	50
H(102)	8862	6496	5522	47	H3B8	7046	5504	8836	50
H(6)	4381	7679	9473	94	H4A8	5609	5615	7406	63
H(10A)	4693	7928	11178	102	H4B8	6074	5302	7711	63
H(10B)	3641	7840	10749	102	H19	6274	6323	8885	35
H(10C)	4877	8253	9765	148	H29	5202	6558	7332	35
H(10D)	3765	8190	9491	148	H39	4675	6601	9318	47
H(10E)	4385	8478	11227	159	H5A9	4396	6114	8826	57
H(10F)	3277	8431	10888	159	H5B9	3396	6276	9038	57
H(10G)	3659	8906	10501	243	H6A9	3092	6322	7373	88
H(10H)	3518	8727	9530	243	H6B9	3206	5983	7692	88
H(10I)	5331	8950	10188	410	H6C9	4079	6148	7179	88
H(10J)	5142	8787	9188	410	H7A9	3662	6821	7650	67
HH2	4561	9310	7800	227	H7B9	4127	7016	8507	67
H2A1	6872	7346	11484	35	H7C9	3196	6814	8694	67
H2B1	6263	7354	10484	35	H110	6067	6946	6845	35
H11	7401	7067	9815	35	H2A10	6330	7424	7970	39
H22	6173	7810	9959	39	H2B10	6702	7397	6893	39
H2A2	7831	8157	9858	40	H111	7711	7673	8353	34
H2B2	6817	8303	10133	40	H2A11	9492	7394	8420	37
H23	6096	8519	7305	49	H2B11	9066	7501	9424	37
H3A3	7173	8958	6746	78	H112	10775	7728	8895	33
H3B3	6463	8991	7643	78	H212	10409	8317	8424	37
H4A3	7920	9110	8340	90	H3A12	11597	8113	9992	49
H4B3	8487	8866	7720	90	H3B12	11786	8434	9548	49
H5A3	8335	8602	9095	54	H512	10672	8834	9828	55
H5B3	7336	8765	9363	54	H612	9398	9036	10719	66
H14	8200	8175	7186	38	H712	8482	8728	11683	66
H24	8116	8145	5188	39	H812	8795	8221	11734	58
H3A4	9891	8225	6251	49	H912	10031	8018	10831	49
H3B4	9854	8134	5140	49	H2A13	14998	7810	8080	58
H44	9137	8678	5884	66	H2B13	14812	7805	6994	58
H5A4	9285	8555	3874	119	H113	11254	7821	7275	34
H5B4	8781	8838	4345	119	H213	12943	8090	6639	36
H5C4	8288	8518	4426	119	H3A13	12645	7462	6932	45
H6A4	10861	8584	4828	114	H3B13	13504	7579	6276	45
H6B4	10837	8654	5957	114	H114	13224	8072	5015	43
H6C4	10495	8900	5192	114	H214	11684	8011	3788	43
H25	8885	7206	6105	36	H314	12762	8072	2514	59
H3A5	9351	7260	4481	51	H4A14	13212	8563	2883	112
H3B5	9899	6983	4970	51	H4B14	13026	8503	3999	112
H3C5	10461	7290	4828	51	H4C14	12140	8494	3235	112
H15	9909	7737	6018	36	H5A14	14127	7826	3087	99
H16	10020	6810	6546	37	H5B14	14312	8074	3892	99
H2A6	11282	6974	8036	38	H5C14	14412	8155	2777	99
H2B6	11570	6714	7315	38	H215	11887	7064	3133	42
H2A7	9320	6686	10800	33	H3A15	10739	7013	1644	49
H2B7	9366	6924	9950	33	H3B15	10306	7100	2673	49
H4A7	7892	6705	9274	32	H4A15	9778	7490	1724	49
H4B7	8007	6447	10060	32	H4B15	10776	7501	1148	49
H57	9451	6277	9361	33	H5A15	11109	7871	2243	48
H17	9313	6768	8148	33	H5B15	10452	7720	3054	48
H18	7870	5975	9283	35	H116	13551	7205	2792	45
H2A8	7379	5802	7489	38	H216	14447	6471	1506	65
H38	3698	5317	8298	97	H2A16	14800	7305	1794	49

Supplementary Table S13b. Hydrogen coordinates (Å x 10^4) and isotropic displacement parameters (Å^2 x 10^3) for the -4 aa truncation of xanthomonin I.

	x	y	z	U(eq)		x	y	z	U(eq)
H2B16	14134	7199	900	49	H3B25	3738	6106	5115	62
H117	7044	5825	2501	34	H3C25	4065	5932	4173	62
H2A17	6485	5521	854	37	H4A25	4131	6845	4114	43
H2B17	5884	5529	1827	37	H4B25	3123	6681	3904	43
H118	5791	5082	2389	39	H5A25	3119	6904	5435	67
H2A18	7461	4747	2741	40	H5B25	4049	6714	5765	67
H2B18	6455	4590	2447	40	H5C25	3035	6552	5552	67
H219	5567	4513	5285	46	H625	5324	6412	5178	31
H3A19	6636	4109	6211	65	H126	6115	5994	5513	31
H3B19	5886	4021	5349	65	H2A26	6106	5537	4231	32
H3A19	7028	4117	5970	65	H2B26	6503	5515	5320	32
H3B19	5890	4050	5816	65	H127	7490	5230	4087	37
H5A19	7783	4289	3680	53	H2A27	9297	5489	3948	35
H5B19	6730	4139	3495	53	H2B27	8825	5352	2980	35
H5A19	7862	4247	4094	53	H128	10530	5134	3633	35
H5B19	7019	4219	3272	53	H228	10121	4565	4282	39
H4AA19	7281	3852	4773	52	H3A28	11351	4708	2714	48
H4AB19	7898	4121	5239	52	H3B28	11403	4389	3220	48
H4BA19	6049	3934	4238	84	H528	9972	4828	1635	58
H4BB19	7094	3826	4641	84	H628	8622	4672	726	76
H120	7850	4755	5385	40	H728	7954	4216	993	76
H220	7895	4877	7337	41	H828	8643	3897	2146	69
H3A20	9538	4736	6215	45	H928	9955	4055	3096	56
H3B20	9597	4819	7332	45	H129	11162	5107	5171	34
H420	8677	4298	6577	75	H2A29	14868	5133	4855	65
H5A20	10034	4040	7204	111	H2B29	14573	5123	5905	65
H5B20	10507	4344	7584	111	H2C29	12390	5451	5709	39
H5C20	10383	4284	6452	111	H2D29	13224	5346	6458	39
H6A20	7944	4416	8004	160	H429	12787	4829	5971	36
H6B20	8969	4466	8543	160	H130	13009	4809	7612	41
H6C20	8644	4140	8219	160	H230	11466	4799	8757	45
H121	9741	5205	6368	39	H330	13482	4844	9344	55
H221	8866	5761	6240	35	H4A30	13201	4403	8542	100
H3A21	10535	5641	7452	56	H4B30	12274	4332	9173	100
H3B21	9490	5733	7827	56	H4C30	13336	4324	9662	100
H3C21	10123	5968	7247	56	H5A30	12426	5005	10591	97
H122	9846	6124	5599	36	H5B30	13083	4726	10901	97
H2A22	10998	5949	4102	37	H5C30	11967	4679	10600	97
H2B22	11408	6198	4826	37	H231	10945	5737	9037	54
H123	9073	6143	3999	33	H3A31	9585	5794	9857	79
H223	9141	6638	2795	32	H3B31	10017	5628	10796	79
H3A23	7552	6222	2961	34	H3A31	9921	5770	10602	79
H3B23	7640	6466	2133	34	H3B31	9456	5663	9586	79
H4A23	8870	6208	1418	35	H5A31	10409	5007	10228	70
H4B23	8974	5986	2310	35	H5B31	10180	4982	9094	70
H324	7126	8230	2823	123	H5A31	10465	4928	9912	70
H124	7622	6955	2875	32	H5B31	9857	5098	9059	70
H224	7356	7154	4662	33	H6A31	8985	5235	10234	47
H3A24	5989	7424	3971	39	H6B31	9056	5338	9122	47
H3B24	6210	7267	2965	39	H4AA31	9161	5313	10531	56
H4A24	7624	7625	3997	47	H4AB31	10212	5335	11051	56
H4B24	7773	7487	2950	47	H132	11716	6119	9629	47
H125	6061	6618	3446	30	H232	11646	6516	12797	59
H225	4299	6375	3350	36	H2A32	12595	6411	10614	48
H3A25	3066	6113	4159	62	H2B32	12890	6123	11231	48

Supplementary Table S14a. Torsion angles [°] for the -4 aa truncation of xanthomonin I.

OH1-C(100)-C(101)-C(102)	-170.7(10)	C47-C57-C67-O37	110.0(4)	O112-C112-N113-C213	6.4(7)
C(100)-C(101)-C(102)-C(103)	175.8(13)	N17-C57-C67-N18	175.6(4)	C212-C112-N113-C213	-174.6(4)
C(101)-C(102)-C(103)-C(104)	-65(3)	C47-C57-C67-N18	-64.6(5)	C112-N113-C213-C113	-126.3(4)
C(102)-C(103)-C(104)-C(105)	-176(2)	O16-C16-N17-C57	-2.5(7)	C112-N113-C213-C313	113.5(5)
C(103)-C(104)-C(105)-OH2	-91(4)	C26-C16-N17-C57	176.9(4)	O113-C113-C213-N113	-32.3(6)
O11-C11-C21-N11	-53.1(5)	C67-C57-N17-C16	-109.8(4)	N114-C113-C213-N113	151.6(4)
N22-C11-C21-N11	126.3(4)	C47-C57-N17-C16	133.2(4)	O113-C113-C213-C313	89.0(5)
C11-C21-N11-C17	88.9(5)	O37-C67-N18-C28	-3.1(6)	N114-C113-C213-C313	-87.1(5)
O11-C11-N22-C22	2.8(7)	C57-C67-N18-C28	171.6(4)	N113-C213-C313-C413	-73.3(5)
C21-C11-N22-C22	-176.5(4)	C67-N18-C28-C38	176.5(4)	C113-C213-C313-C413	164.9(4)
C11-N22-C22-C12	127.6(4)	C67-N18-C28-C18	-60.0(5)	C213-C313-C413-O313	86.9(6)
O12-C12-C22-N22	1.4(6)	O18-C18-C28-N18	143.2(4)	C213-C313-C413-N213	-91.0(5)
N13-C12-C22-N22	-179.8(4)	N19-C18-C28-N18	-40.7(5)	O113-C113-N114-C214	-2.2(7)
O12-C12-N13-C53	-178.8(4)	O18-C18-C28-C38	-95.0(5)	C213-C113-N114-C214	173.9(4)
C22-C12-N13-C53	2.4(6)	N19-C18-C28-C38	81.1(5)	C113-N114-C214-C114	-69.3(5)
O12-C12-N13-C23	10.8(7)	N18-C28-C38-C48	-177.8(4)	C113-N114-C214-C314	167.8(4)
C22-C12-N13-C23	-168.0(4)	C18-C28-C38-C48	59.4(6)	O114-C114-C214-N114	-33.8(6)
C12-N13-C23-C13	-77.8(6)	C28-C38-C48-C58	-140.6(5)	N115-C114-C214-N114	151.1(4)
C53-N13-C23-C13	110.7(5)	C38-C48-C58-O28	-17.1(9)	O114-C114-C214-C314	89.9(6)
C12-N13-C23-C33	160.8(5)	C38-C48-C58-O38	162.4(5)	N115-C114-C214-C314	-85.2(5)
C53-N13-C23-C33	-10.7(5)	O18-C18-N19-C29	-6.0(7)	N114-C214-C314-C514	60.6(6)
O13-C13-C23-N13	175.5(5)	C28-C18-N19-C29	178.1(4)	C114-C214-C314-C514	-62.5(6)
N14-C13-C23-N13	-5.5(7)	C18-N19-C29-C19	-118.8(4)	N114-C214-C314-C414	-66.4(7)
O13-C13-C23-C33	-67.8(7)	C18-N19-C29-C39	120.2(5)	C114-C214-C314-C414	170.4(6)
N14-C13-C23-C33	111.2(6)	O19-C19-C29-N19	-70.7(5)	O114-C114-N115-C515	170.0(4)
N13-C23-C33-C43	29.4(6)	N110-C19-C29-N19	108.5(4)	C214-C114-N115-C515	-14.8(7)
C13-C23-C33-C43	-96.6(6)	O19-C19-C29-C39	51.3(5)	O114-C114-N115-C215	2.1(7)
C23-C33-C43-C53	-37.9(7)	N110-C19-C29-C39	-129.5(4)	C214-C114-N115-C215	177.2(4)
C12-N13-C53-C43	176.9(5)	N19-C29-C39-C79	179.2(4)	C114-N115-C215-C315	156.5(4)
C23-N13-C53-C43	-11.9(6)	C19-C29-C39-C79	60.0(5)	C515-N115-C215-C315	-13.0(5)
C33-C43-C53-N13	30.4(7)	N19-C29-C39-C59	-54.1(5)	C114-N115-C215-C115	-82.7(5)
O13-C13-N14-C24	3.1(8)	C19-C29-C39-C59	-173.3(4)	C515-N115-C215-C115	107.7(4)
C23-C13-N14-C24	-175.9(4)	C79-C39-C59-C69	62.1(6)	O115-C115-C215-N115	-112.3(5)
C13-N14-C24-C34	104.9(5)	C29-C39-C59-C69	-63.3(6)	N116-C115-C215-N115	68.3(5)
C13-N14-C24-C14	-126.4(5)	O19-C19-N110-C210	-2.7(6)	O115-C115-C215-C315	2.3(7)
O14-C14-C24-N14	63.3(5)	C29-C19-N110-C210	178.1(4)	N116-C115-C215-C315	-177.2(4)
N15-C14-C24-N14	-118.6(4)	C19-N110-C210-C110	70.7(5)	N115-C215-C315-C415	31.7(5)
O14-C14-C24-C34	-169.9(4)	O110-C110-C210-N110	23.3(6)	C115-C215-C315-C415	-87.2(5)
N15-C14-C24-C34	8.2(6)	N111-C110-C210-N110	-157.8(4)	C215-C315-C415-C515	-39.3(5)
N14-C24-C34-C44	-65.8(6)	O110-C110-N111-C211	-0.2(6)	C114-N115-C515-C415	-179.7(4)
C14-C24-C34-C44	169.2(4)	C210-C110-N111-C211	-179.1(4)	C215-N115-C515-C415	-10.7(5)
C24-C34-C44-C64	175.7(5)	C110-N111-C211-C111	144.2(4)	C315-C415-C515-N115	30.4(5)
C24-C34-C44-C54	-59.0(7)	O111-C111-C211-N111	14.4(6)	O115-C115-N116-C216	3.3(8)
O15-C15-C25-N15	-7.1(6)	N112-C111-C211-N111	-167.8(4)	C215-C115-N116-C216	-177.2(4)
N16-C15-C25-N15	172.0(4)	O111-C111-N112-C212	1.3(6)	C115-N116-C216-C116	-100.0(6)
O15-C15-C25-C35	113.9(5)	C211-C111-N112-C212	-176.5(4)	O116-C116-C216-N116	-130.1(5)
N16-C15-C25-C35	-67.0(5)	C111-N112-C212-C112	-120.5(4)	O216-C116-C216-N116	49.9(6)
O14-C14-N15-C25	4.6(6)	C111-N112-C212-C312	113.2(4)	C523-N117-C217-C117	92.9(5)
C24-C14-N15-C25	-173.3(4)	O112-C112-C212-N112	-165.3(4)	O117-C117-C217-N117	-52.8(6)
C35-C25-N15-C14	100.9(5)	N113-C112-C212-N112	15.7(6)	N118-C117-C217-N117	126.6(4)
C15-C25-N15-C14	-137.4(4)	O112-C112-C212-C312	-39.0(6)	O117-C117-N118-C218	6.3(7)
O15-C15-N16-C26	1.0(7)	N113-C112-C212-C312	141.9(4)	C217-C117-N118-C218	-173.0(4)
C25-C15-N16-C26	-178.0(4)	N112-C212-C312-C412	-77.6(5)	C117-N118-C218-C118	126.1(4)
C15-N16-C26-C16	129.9(5)	C112-C212-C312-C412	155.0(4)	O118-C118-C218-N118	-4.6(6)
O16-C16-C26-N16	151.4(4)	C212-C312-C412-C512	-92.5(6)	N119-C118-C218-N118	176.6(4)
N17-C16-C26-N16	-28.0(6)	C212-C312-C412-C912	85.2(6)	O118-C118-N119-C519	180.0(4)
C21-N11-C17-O17	9.0(6)	C912-C412-C512-C612	0.4(8)	C218-C118-N119-C519	-1.3(7)
C21-N11-C17-C27	-170.7(4)	C312-C412-C512-C612	178.3(5)	O118-C118-N119-C219	11.8(6)
O17-C17-C27-C47	137.3(4)	C412-C512-C612-C712	0.6(9)	C218-C118-N119-C219	-169.4(4)
N11-C17-C27-C47	-43.1(5)	C512-C612-C712-C812	-0.9(9)	C118-N119-C219-C119	-74.5(6)
C17-C27-C47-C57	170.8(3)	C612-C712-C812-C912	0.2(9)	C519-N119-C219-C119	115.9(5)
C27-C47-C57-N17	-55.6(5)	C712-C812-C912-C412	0.8(8)	C118-N119-C219-C319	165.5(4)
C27-C47-C57-C67	-174.5(4)	C512-C412-C912-C812	-1.2(7)	C519-N119-C219-C319	-4.0(5)
N17-C57-C67-O37	-9.7(6)	C312-C412-C912-C812	-178.9(5)	O119-C119-C219-N119	164.2(5)

Supplementary Table S14b. Torsion angles [°] for the -4 aa truncation of xanthomonin I.

N120-C119-C219-N119	-21.4(7)	C323-C423-C523-N117	-44.9(5)	O128-C128-N129-C429	3.7(7)
O119-C119-C219-C319	-79.9(6)	O123-C123-N124-C224	1.4(6)	C228-C128-N129-C429	-179.9(4)
N120-C119-C219-C319	94.5(5)	C223-C123-N124-C224	175.4(3)	C429-C229-C329-O229	90.4(6)
N119-C219-C319-C4A19	24.1(7)	C123-N124-C224-C324	172.8(4)	C429-C229-C329-N229	-88.2(5)
C119-C219-C319-C4A19	-99.8(7)	C123-N124-C224-C124	-61.1(5)	C128-N129-C429-C229	130.4(4)
N119-C219-C319-C4B19	-17.1(9)	O124-C124-C224-N124	151.5(4)	C128-N129-C429-C129	-108.9(5)
C119-C219-C319-C4B19	-141.1(9)	N125-C124-C224-N124	-28.9(5)	C329-C229-C429-N129	-76.9(5)
C118-N119-C519-C4A19	174.7(6)	O124-C124-C224-C324	-84.0(5)	C329-C229-C429-C129	161.5(4)
C219-N119-C519-C4A19	-16.5(7)	N125-C124-C224-C324	95.6(4)	O129-C129-C429-N129	-34.0(6)
C118-N119-C519-C4B19	-146.0(8)	N124-C224-C324-C424	-69.0(5)	N130-C129-C429-N129	148.3(4)
C219-N119-C519-C4B19	22.8(8)	C124-C224-C324-C424	165.5(4)	O129-C129-C429-C229	87.6(5)
C4B19-C319-C4A19-C519	62.8(10)	C224-C324-C424-C524	-174.9(4)	N130-C129-C429-C229	-90.1(5)
C219-C319-C4A19-C519	-36.2(9)	C324-C424-C524-O224	-28.5(7)	O129-C129-N130-C230	-2.3(7)
N119-C519-C4A19-C319	32.5(9)	C324-C424-C524-O324	154.9(5)	C429-C129-N130-C230	175.4(4)
C4B19-C519-C4A19-C319	-63.9(12)	O124-C124-N125-C625	-12.8(6)	C129-N130-C230-C130	-59.9(5)
C4A19-C319-C4B19-C519	-60.8(10)	C224-C124-N125-C625	167.6(4)	C129-N130-C230-C330	176.9(4)
C219-C319-C4B19-C519	31.4(12)	C325-C225-C425-C525	53.6(6)	O130-C130-C230-N130	-47.7(6)
N119-C519-C4B19-C319	-32.4(11)	C625-C225-C425-C525	-72.0(5)	N131-C130-C230-N130	134.6(4)
C4A19-C519-C4B19-C319	58.7(10)	C124-N125-C625-C125	-130.6(4)	O130-C130-C230-C330	74.5(6)
O119-C119-N120-C220	6.8(8)	C124-N125-C625-C225	108.6(4)	N131-C130-C230-C330	-103.1(5)
C219-C119-N120-C220	-167.3(4)	O125-C125-C625-N125	-62.0(5)	N130-C230-C330-C430	-62.3(6)
C119-N120-C220-C320	120.5(5)	N126-C125-C625-N125	119.6(4)	C130-C230-C330-C430	176.2(5)
C119-N120-C220-C120	-111.5(5)	O125-C125-C625-C225	60.4(5)	N130-C230-C330-C530	173.6(5)
O120-C120-C220-N120	59.1(5)	N126-C125-C625-C225	-118.0(4)	C130-C230-C330-C530	52.1(6)
N121-C120-C220-N120	-120.9(4)	C425-C225-C625-N125	-60.9(5)	O130-C130-N131-C531	175.2(5)
O120-C120-C220-C320	-175.6(4)	C325-C225-C625-N125	172.6(4)	C230-C130-N131-C531	-7.1(7)
N121-C120-C220-C320	4.4(6)	C425-C225-C625-C125	179.9(4)	O130-C130-N131-C231	5.5(7)
N120-C220-C320-C420	-59.0(6)	C325-C225-C625-C125	53.4(5)	C230-C130-N131-C231	-176.8(4)
C120-C220-C320-C420	175.8(4)	O125-C125-N126-C226	-5.3(6)	C130-N131-C231-C131	-64.3(6)
C220-C320-C420-C620	-61.7(7)	C625-C125-N126-C226	173.0(4)	C531-N131-C231-C131	124.6(5)
C220-C320-C420-C520	175.2(5)	C125-N126-C226-C126	78.6(5)	C130-N131-C231-C331	176.2(5)
O120-C120-N121-C221	8.4(7)	O126-C126-C226-N126	7.7(6)	C531-N131-C231-C331	5.1(6)
C220-C120-N121-C221	-171.6(4)	N127-C126-C226-N126	-170.9(3)	O131-C131-C231-N131	-37.6(7)
C120-N121-C221-C121	-134.4(4)	O126-C126-N127-C227	0.9(7)	N132-C131-C231-N131	147.0(4)
C120-N121-C221-C321	104.0(5)	C226-C126-N127-C227	179.6(4)	O131-C131-C231-C331	78.1(7)
O121-C121-C221-N121	-17.5(5)	C126-N127-C227-C127	144.6(4)	N132-C131-C231-C331	-97.2(6)
N122-C121-C221-N121	164.4(4)	O127-C127-C227-N127	14.4(6)	N131-C231-C331-C4A31	24.3(7)
O121-C121-C221-C321	104.6(5)	N128-C127-C227-N127	-167.1(4)	C131-C231-C331-C4A31	-98.1(7)
N122-C121-C221-C321	-73.5(5)	O127-C127-N128-C228	2.2(6)	N131-C231-C331-C631	-21.0(8)
O121-C121-N122-C222	0.9(6)	C227-C127-N128-C228	-176.2(4)	C131-C231-C331-C631	-143.4(7)
C221-C121-N122-C222	178.9(4)	C127-N128-C228-C128	-121.0(4)	C130-N131-C531-C631	-155.6(7)
C121-N122-C222-C122	139.7(4)	C127-N128-C228-C328	113.4(5)	C231-N131-C531-C631	14.7(8)
O122-C122-C222-N122	142.5(4)	O128-C128-C228-N128	-172.7(4)	C130-N131-C531-C4A31	164.5(6)
N123-C122-C222-N122	-38.4(5)	N129-C128-C228-N128	10.7(6)	C231-N131-C531-C4A31	-25.1(6)
O122-C122-N123-C223	0.5(6)	O128-C128-C228-C328	-47.4(6)	N131-C531-C631-C331	-25.5(9)
C222-C122-N123-C223	-178.6(4)	N129-C128-C228-C328	136.0(4)	C4A31-C531-C631-C331	47.3(7)
C122-N123-C223-C123	-102.2(4)	N128-C228-C328-C428	-71.4(5)	C4A31-C331-C631-C531	-65.3(8)
C122-N123-C223-C323	140.9(4)	C128-C228-C328-C428	161.4(4)	C231-C331-C631-C531	28.6(9)
O123-C123-C223-N123	-12.9(5)	C228-C328-C428-C528	92.6(6)	C231-C331-C4A31-C531	-41.6(8)
N124-C123-C223-N123	173.4(3)	C228-C328-C428-C928	-83.0(6)	C631-C331-C4A31-C531	57.8(8)
O123-C123-C223-C323	105.7(4)	C928-C428-C528-C628	-1.3(8)	N131-C531-C4A31-C331	42.1(8)
N124-C123-C223-C323	-68.1(4)	C328-C428-C528-C628	-177.1(5)	C631-C531-C4A31-C331	-73.8(10)
N123-C223-C323-C423	-54.6(5)	C428-C528-C628-C728	0.4(9)	O131-C131-N132-C232	-12.3(7)
C123-C223-C323-C423	-172.3(4)	C528-C628-C728-C828	-0.7(10)	C231-C131-N132-C232	162.9(4)
C223-C323-C423-C523	169.4(4)	C628-C728-C828-C928	1.8(10)	C131-N132-C232-C132	-76.5(6)
C217-N117-C523-O223	6.2(6)	C728-C828-C928-C428	-2.7(9)	O132-C132-C232-N132	-9.9(8)
C217-N117-C523-C423	-171.0(4)	C528-C428-C928-C828	2.4(7)	O232-C132-C232-N132	172.2(4)
C323-C423-C523-O223	137.9(4)	C328-C428-C928-C828	178.3(5)		

Supplementary Table S15. Hydrogen bonds for the -4 aa truncation of xanthomonin I [Å and °].

D-H...A	d(D-H)	d(H...A)	d(D...A)	<(DHA)
O(100)-H(101)...O37	0.91	1.68	2.582(4)	171.1
O(100)-H(102)...O123	0.98	1.66	2.608(4)	162.2
O(100)-H(102)...N123	0.98	2.69	3.292(5)	120.2
OH1-H(6)...N213#1	0.84	2.69	3.347(7)	136.3
OH1-H(6)...O313#1	0.84	1.87	2.700(6)	167.8
OH2-HH2...OW11	0.84	2.14	2.777(13)	132.6
OH2-HH2...OW21	0.84	2.07	2.64(2)	125.2
N11-H11...O19	0.88	2.10	2.800(5)	135.5
N22-H22...OH1	0.88	2.07	2.936(6)	169.1
N14-H14...O111	0.88	1.91	2.770(5)	164.8
N15-H15...O113	0.88	2.14	2.858(5)	138.3
N16-H16...O(100)	0.88	1.88	2.740(5)	164.8
N17-H17...O110	0.88	2.01	2.834(5)	156.5
N18-H18...O223#2	0.88	1.91	2.782(5)	172.9
O38-H38...O130#1	0.84	1.79	2.561(6)	151.4
O38-H38...N130#1	0.84	2.69	3.199(6)	120.0
N19-H19...OW1	0.88	2.10	2.949(5)	160.8
N110-H110...O124	0.88	2.09	2.932(5)	159.8
N111-H111...O12	0.88	2.32	3.178(5)	165.2
N112-H112...OW8	0.88	1.98	2.854(5)	171.2
N213-H2A13...O12#3	0.88	2.16	2.837(5)	132.8
N213-H2B13...OW3	0.88	2.18	3.027(7)	161.8
N113-H113...O15	0.88	1.99	2.800(5)	152.5
N114-H114...OW2	0.88	2.16	3.028(6)	169.0
N116-H116...O114	0.88	2.10	2.852(5)	142.7
O216-H216...OW6	0.84	1.81	2.648(5)	172.7
N117-H117...O125	0.88	2.18	2.915(5)	140.2
N118-H118...OW5	0.88	2.04	2.920(5)	174.4
N120-H120...O127	0.88	1.97	2.809(5)	158.7
N121-H121...O129	0.88	2.19	2.894(5)	136.9
N122-H122...O(100)	0.88	1.88	2.729(5)	161.6
N123-H123...O126	0.88	1.96	2.793(5)	158.1
O324-H324...OW14	0.84	1.97	2.778(14)	161.4
O324-H324...OW16	0.84	1.87	2.638(14)	151.7
O324-H324...OW30	0.84	2.41	3.24(4)	170.7
N124-H124...O17#4	0.88	1.99	2.777(5)	148.7
N125-H125...O116#1	0.88	2.53	3.270(5)	142.7
N126-H126...O18	0.88	1.92	2.798(5)	175.4
N127-H127...O118	0.88	2.28	3.120(5)	158.9
N128-H128...OW15	0.88	2.03	2.894(6)	165.2
N129-H129...O121	0.88	1.98	2.818(5)	159.4
N229-H2A29...O118#3	0.88	2.04	2.878(6)	159.5
N130-H130...OW23	0.88	2.12	2.988(7)	167.2
N132-H132...O16	0.88	2.01	2.844(5)	159.0
O232-H232...O122#2	0.84	1.78	2.613(5)	169.0

Supporting Information References

[1] J. D. Hegemann, M. Zimmermann, X. Xie, M. A. Marahiel, *J Am Chem Soc* **2013**, *135*, 210-222.

[2] J. Chiu, P. E. March, R. Lee, D. Tillett, *Nucleic Acids Res* **2004**, *32*, e174.

[3] J. Chiu, D. Tillett, I. W. Dawes, P. E. March, *J Microbiol Methods* **2008**, *73*, 195-198.

[4] U. Mueller, N. Darowski, M. R. Fuchs, R. Forster, M. Hellmig, K. S. Paithankar, S. Puhringer, M. Steffien, G. Zocher, M. S. Weiss, *J Synchrotron Radiat* **2012**, *19*, 442-449.

[5] W. Kabsch, *Acta Crystallogr D Biol Crystallogr* **2010**, *66*, 125-132.

[6] G. M. Sheldrick, *Acta Crystallogr A* **2008**, *64*, 112-122.

[7] G. M. Sheldrick, University of Göttingen, **2013**.

[8] M. Rance, O. W. Sorensen, G. Bodenhausen, G. Wagner, R. R. Ernst, K. Wuthrich, *Biochem Biophys Res Commun* **1983**, *117*, 479-485.

[9] A. Bax, D. G. Davis, *Journal of Magnetic Resonance* **1985**, *65*, 355-360.

[10] J. Jeener, B. H. Meier, P. Bachmann, R. R. Ernst, *Journal of Chemical Physics* **1979**, *71*, 4546-4553.

[11] K. Wüthrich, *NMR of proteins and nucleic acids*, Wiley, New York, **1986**.

[12] T. D. Goddard, D. J. Kneller, University of California, San Fransisco, **2007**.

[13] T. Herrmann, P. Guntert, K. Wuthrich, *J Mol Biol* **2002**, *319*, 209-227.

[14] J. D. Hegemann, M. Zimmermann, S. Zhu, D. Klug, M. A. Marahiel, *Biopolymers* **2013**.

[15] M. Zimmermann, J. D. Hegemann, X. Xie, M. A. Marahiel, *Chem Biol* **2013**, *20*, 558-569.

[16] Tripos, 7.3 ed., Tripos Inc. 1699 South Hanley Rd. St. Louis, MO 631444, **2006**.

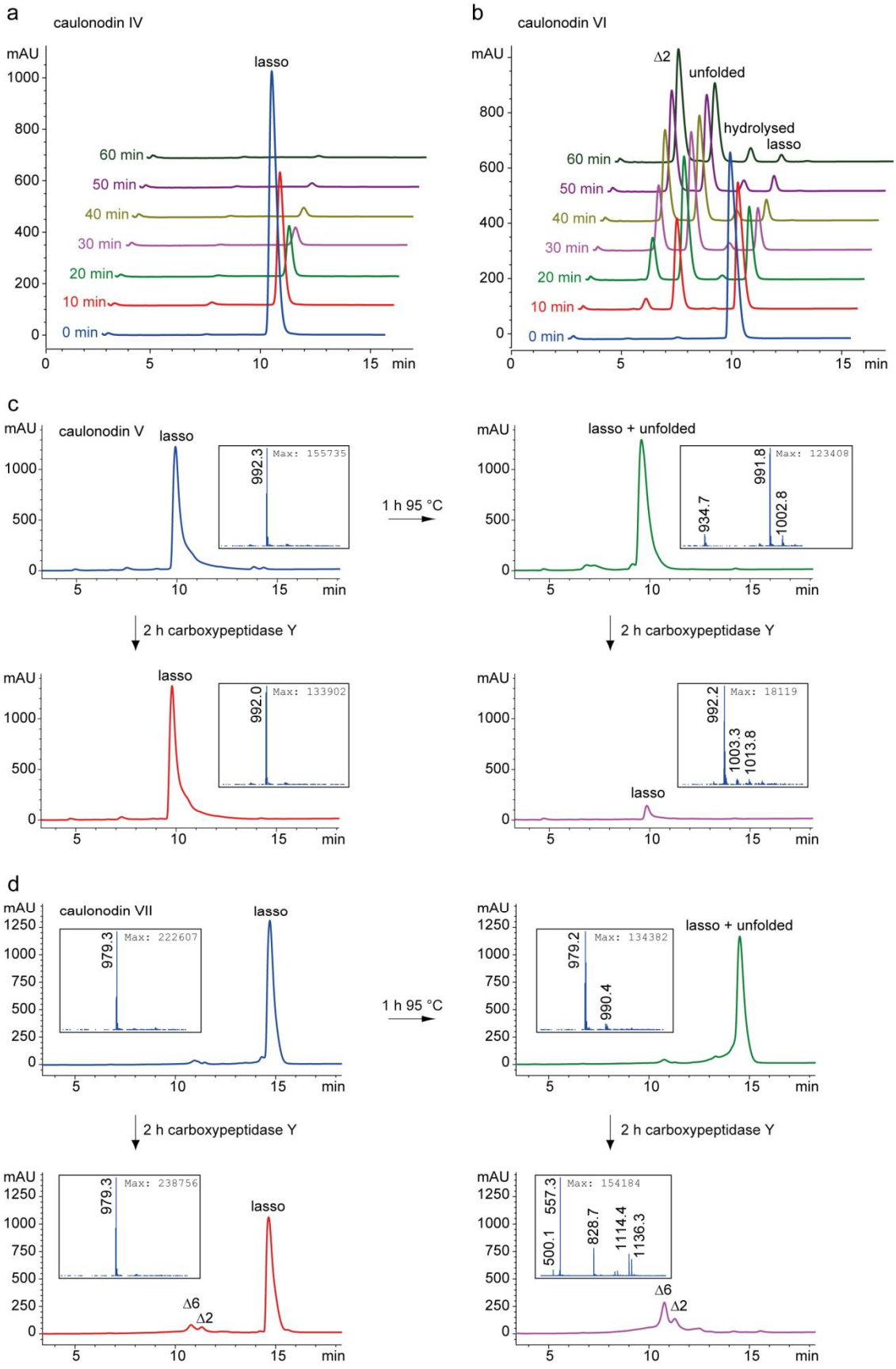
Characterization of Caulonodin Lasso Peptides Revealed Unprecedented N-Terminal Residues and a Precursor Motif Essential for Peptide Maturation

*Marcel Zimmermann, Julian D. Hegemann, Xiulan Xie and Mohamed A. Marahiel**

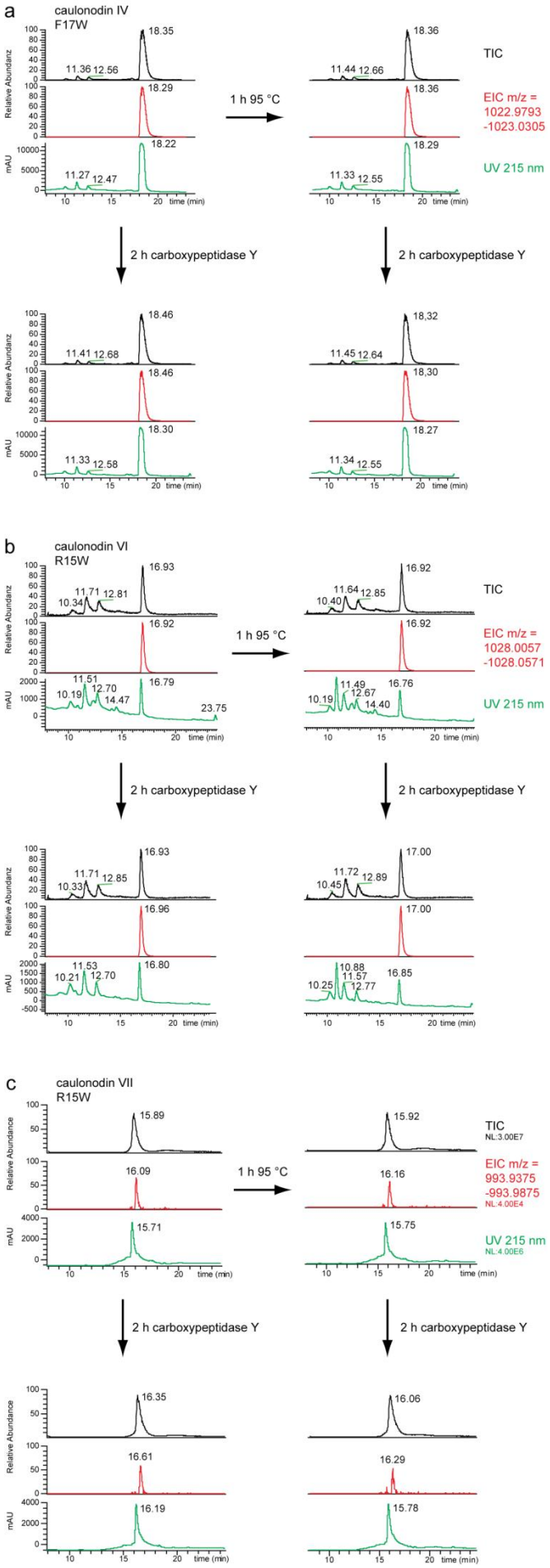
Department of Chemistry, Biochemistry, Philipps-University Marburg, Hans-Meerwein-Strasse 4 and LOEWE-Center for Synthetic Microbiology, D-35032, Marburg, Germany

*marahiel@staff.uni-marburg.de

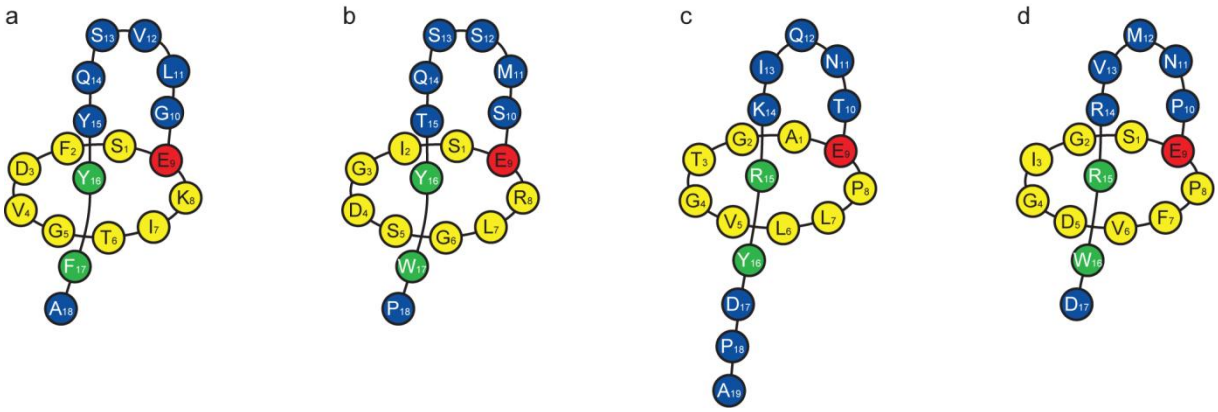
Supporting Information



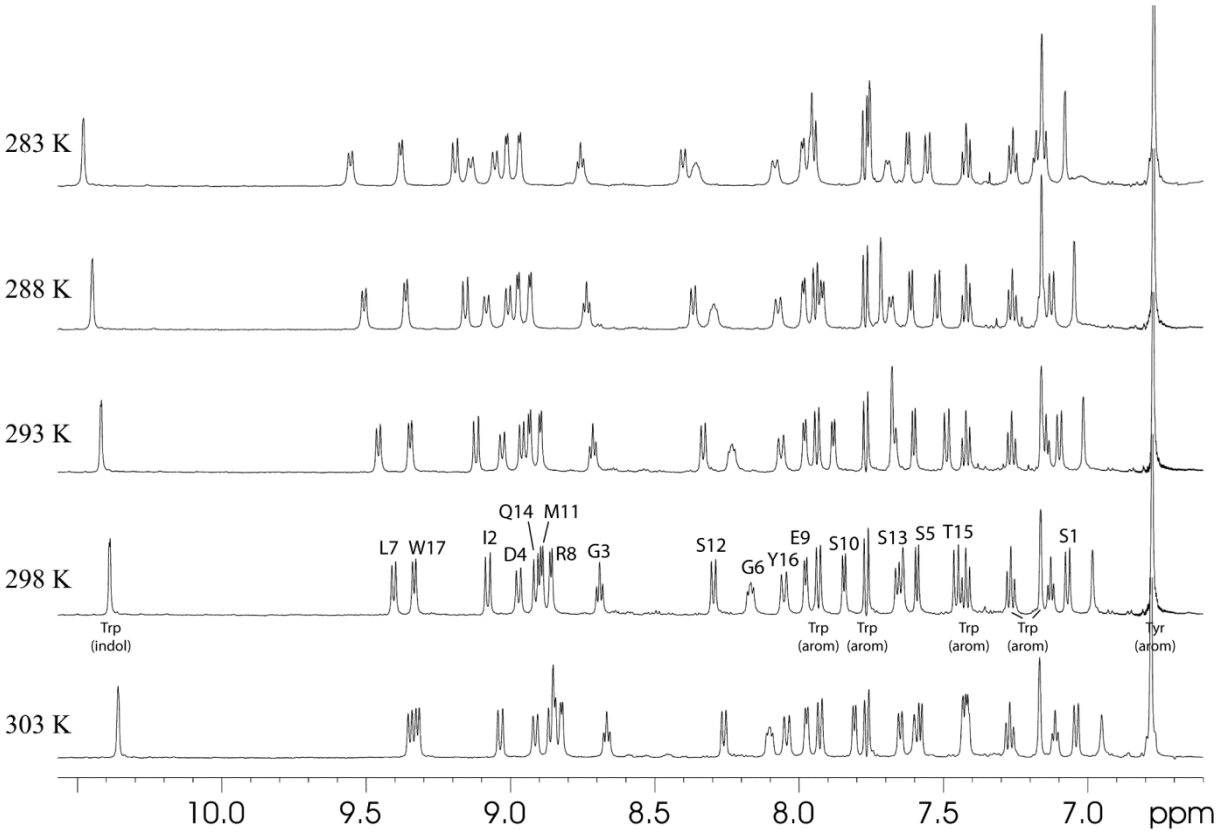
Supporting figure 1 – heat stability of **(a)** caulonodin IV and **(b)** caulonodin VI, heat and carboxypeptidase stability of **(c)** caulonodin V and **(d)** caulonodin VII shown as UV traces and mass signals measured on the low-resolution HPLC-MS system



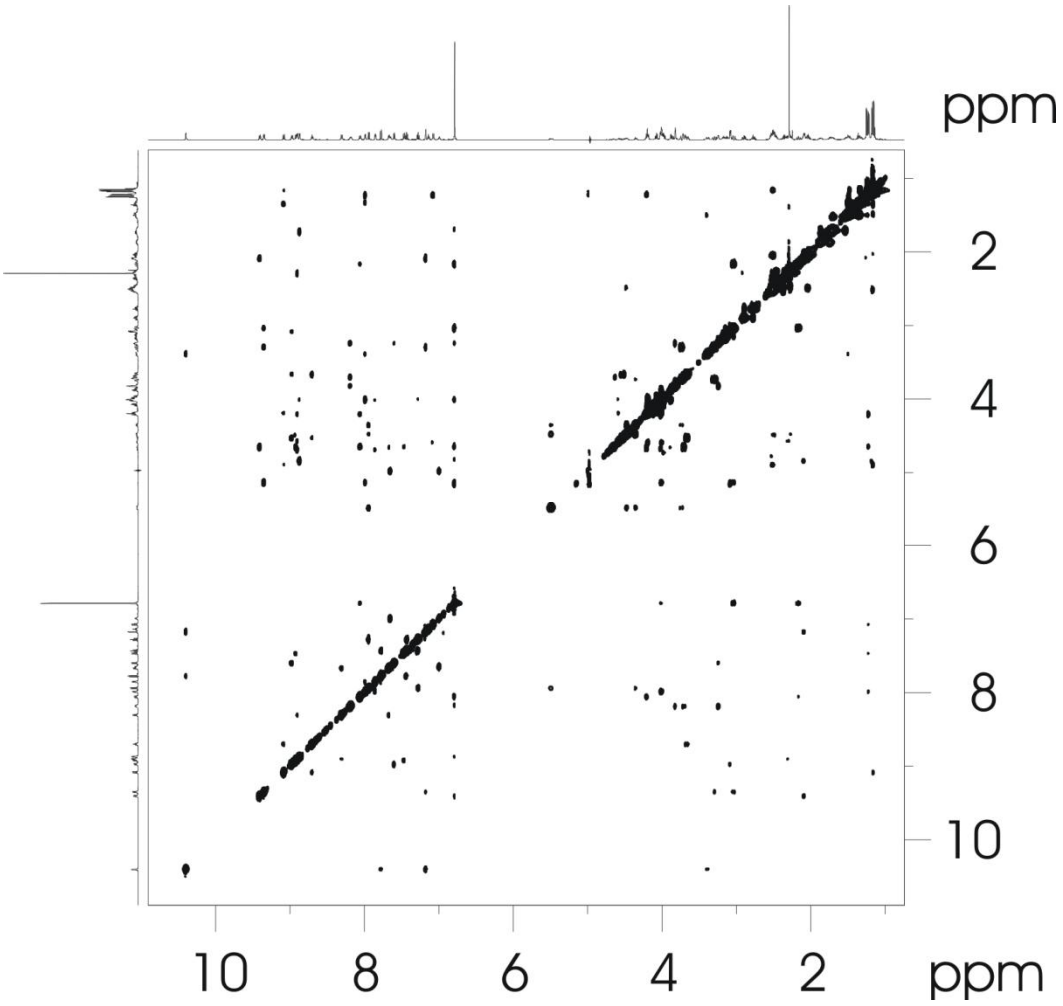
Supporting figure 2 – heat stable mutants of **(a)** caulonodin IV, **(b)** caulonodin VI and **(c)** caulonodin VII shown as total ion current (TIC), extracted ion current (EIC) and UV traces measured on the high-resolution HPLC-MS system



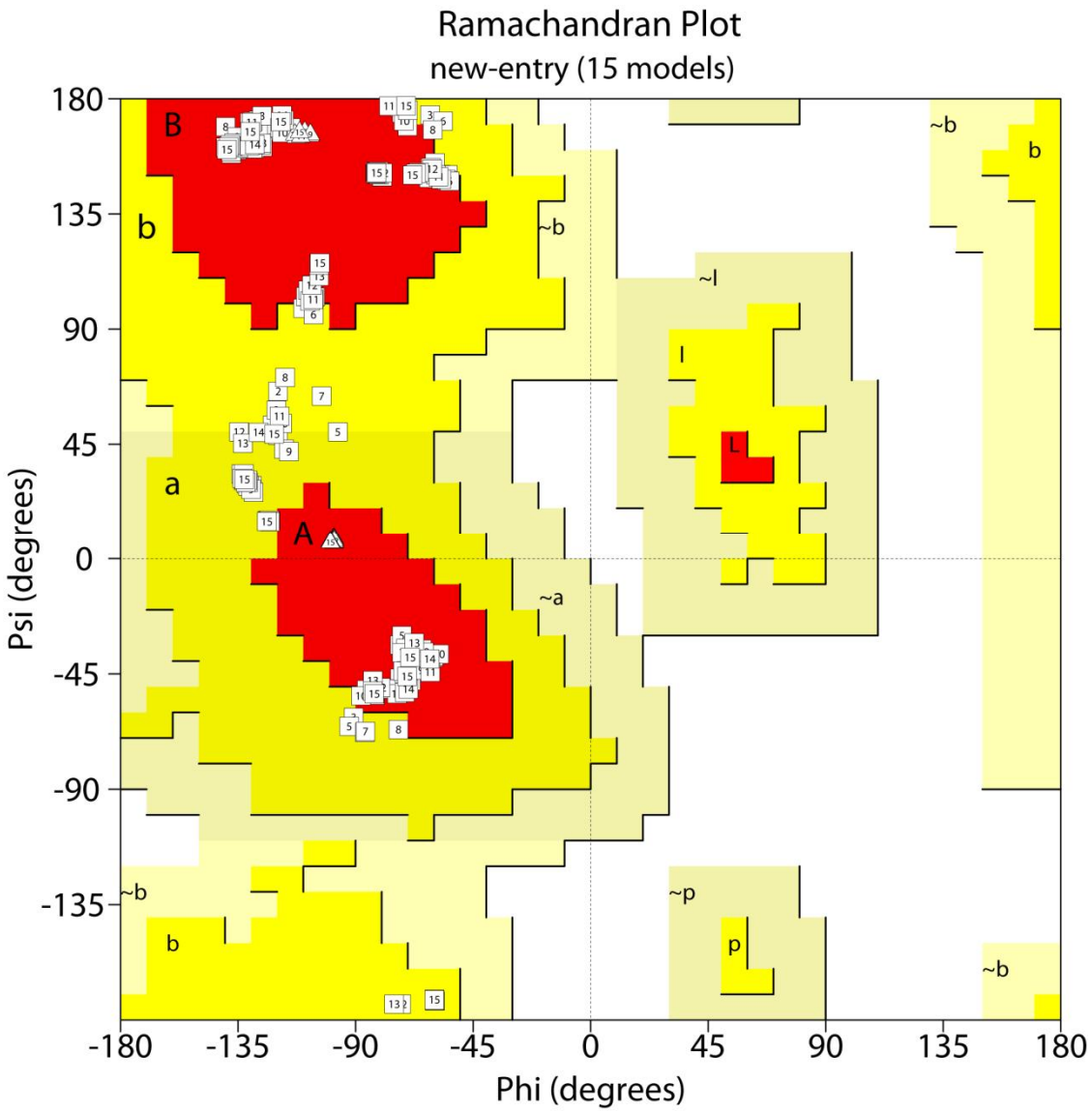
Supporting figure 3 – 3D fold predictions of the caulonodins **(a)** IV, **(b)** V, **(c)** VI and **(d)** VII in a schematic representation based on the plug mutants; aa involved in the formation of the macrolactam ring shown in red, other aa of the ring shown in yellow, aa in the tail shown in blue, plug aa shown in green.



Supporting figure 4 – ^{1}H NMR spectra of caulonodin V in H_2O/D_2O (9:1) at variable temperatures in the range from 10.6 to 6.6 ppm



Supporting figure 5 – 2D ^{1}H NOESY spectrum of caulonodin V at 300 ms mixing time

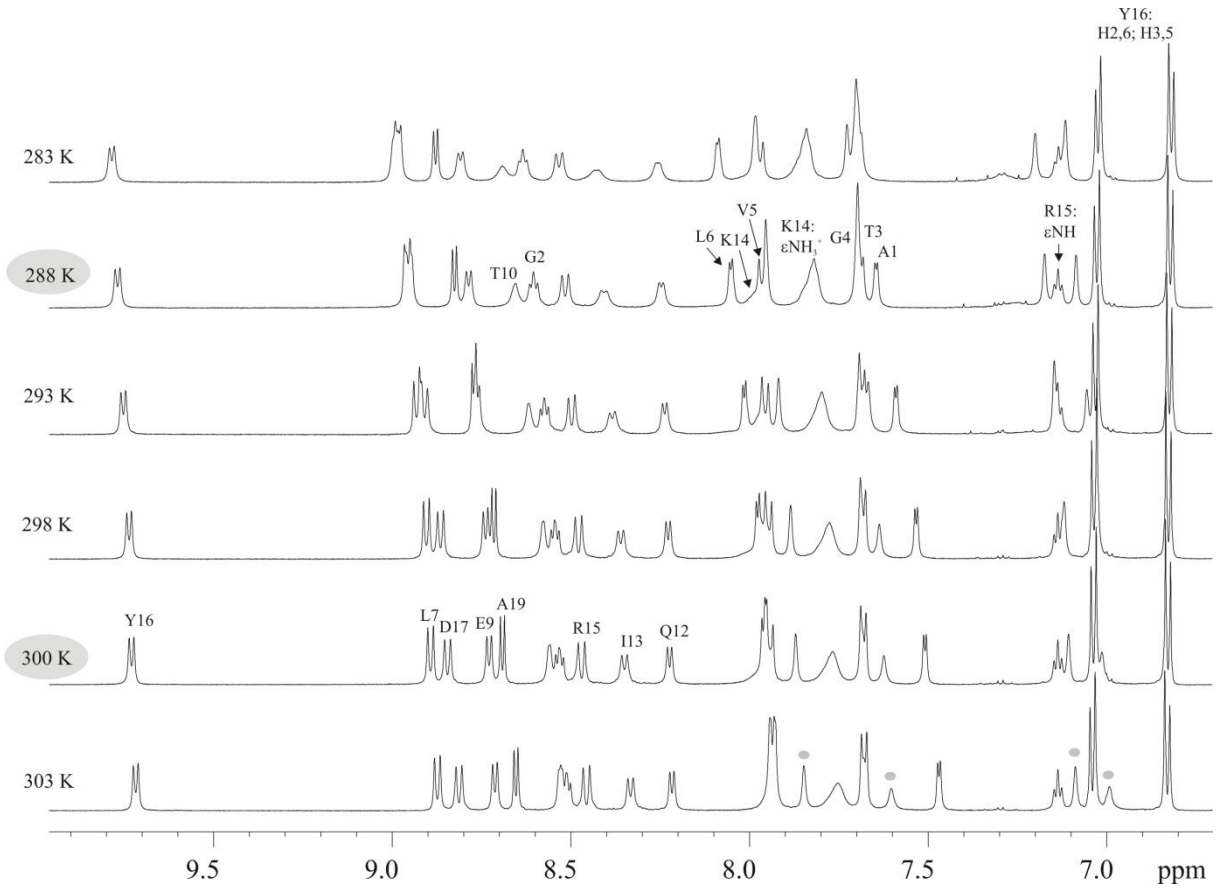


Plot statistics

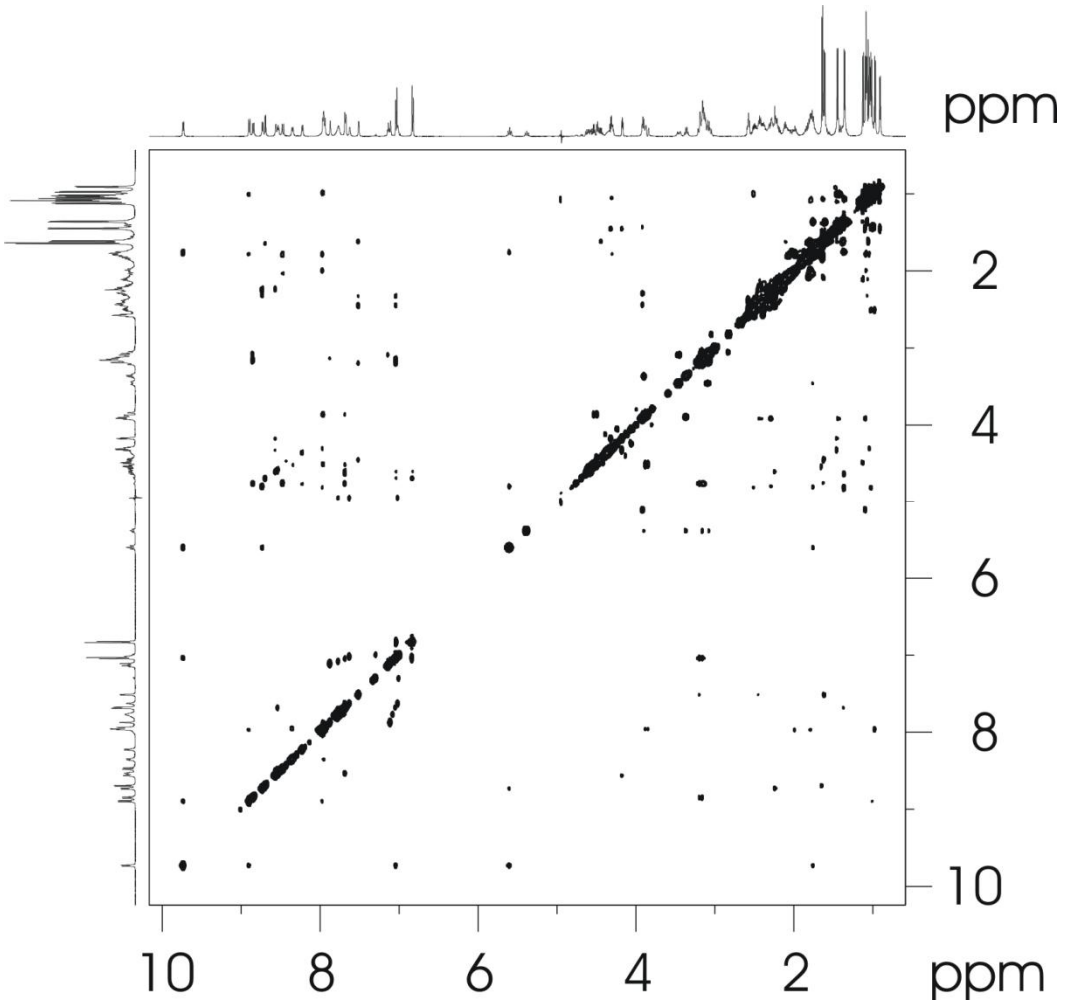
Residues in most favored regions [A,B,L]	132	62.9%
Residues in additional allowed regions [a,b,l,p]	78	37.1%
Residues in generously allowed regions [~a,~b,~l,~p]	0	0.0%
Residues in disallowed regions	0	0.0%
Number of non-glycine and non-proline residues	210	100.0%
Number of end-residues (excl. Gly and Pro)	15	
Number of glycine residues (shown as triangles)	30	
Number of proline residues	15	
Total number of residues	270	

Based on an analysis of 118 structures of resolution of at least 2.0 Angstroms and R-factor no greater than 20%, a good quality model would be expected to have over 90% in the most favoured regions.
Model numbers shown inside each data point.

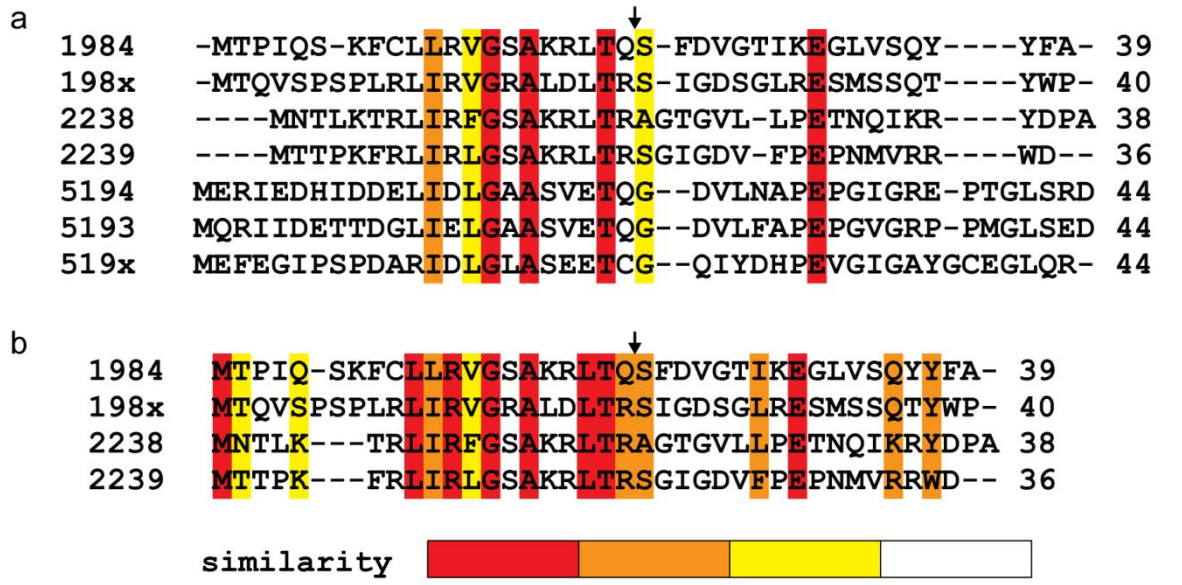
Supporting figure 6 – Quality of the structure determination. Ramachandran Plot of the 15 minimum energy structures showing all residues in most favored and additionally allowed regions.



Supporting figure 7 – ^{1}H NMR spectra of caulonodin VI in H_2O/D_2O (9:1) at variable temperatures in the range from 10.0 to 6.7 ppm; labeling according to signal assignments at 288 and 300 K; circles denote side chain NH2 of N11 and Q12.



Supporting figure 8 – 2D ^{1}H NOESY spectrum of caulonodin VI at 300 ms mixing time



Supporting figure 9 – Multiple alignments using the gonnet protein weight matrix of **(a)** all seven caulonodin precursors and **(b)** the four chromosomal precursors; coloring according to similarity from red for identical and orange to yellow and white for different; arrow marks proteolytic cleavage site between leader and core peptide

Name	Start	*p*-value	Sites
Caulonodin_I	10	8.15e-15	MERIEDHID DELIDLGAASVETQ
Caulonodin_II	10	7.81e-14	MQRIIDETT DGLIELGAASVETQ
Caulonodin_VII	6	2.77e-12	MTTPK FRLIRLGSAKRLTR
Sphingopyxin_II	8	5.99e-12	MERTEVI EEVIDLGKASVETK
Caulosegnin_II	5	1.24e-11	MTKT HRLIRLGDAQRLTQ
Sphingopyxin_I	5	1.43e-11	MKDF NELIDLGAISVETR
Astexin-1	15	1.43e-11	IISETVQPKT AGLIVLGKASAETR
Sphingonodin_I	5	5.42e-11	MERD NDVIELGAVSVETK
Xanthomonin_III	16	8.89e-11	DSAPSQEDHG KDIIVLGIASVETQ
Caulosegnin_I	10	1.61e-10	MTKKNATQA PRLVRVGDAHRLTQ
Astexin-3	12	2.28e-10	TKRTTIAARR VGLIDLGKATRQTK
Caulonodin_IV	8	3.19e-10	MTPIQSK FCLLRVGSAKRLTQ
Syanodin	10	4.43e-10	MERNSEDRR DDVVELGAVSVETK
Sphingonodin_II	10	4.94e-10	MDRHDNSEV DEIIDLGTASAVTQ
Xanthomonin_II	15	5.49e-10	NNDARTTALD QDLIVLGVASLDTQ
Caulonodin_VI	6	5.49e-10	MNTLK TRLIRFGSAKRLTR
Rhodanodin	21	1.03e-09	TNENIRSNAQ DDVIELSVASVETK
Rubrivinodin	12	1.26e-09	KEFAMDEELE LEIVDLGDAKELTQ
Caulosegnin_III	5	1.54e-09	MTSR FQLLRLGKADRLTR
SSV-2083	23	1.87e-09	MTSTDELYEA PELIEIGDYAELTR
Caulonodin_III	10	2.51e-09	MEFEGIPSP DARIDLGLASEETC
Xanthomonin_I	13	2.16e-08	SNDTTHSDAS NEITVLGVASTDTK
Caulonodin_V	9	6.72e-08	MTQVSPSP LRLIRVGRALDLTR
Astexin-2	12	1.81e-07	RTYNRSLPAR AGLTDLGKVTTHTK
Zucinodin	10	4.61e-07	MTRLLNLMS VRLLGFGSAKAATN
Lariatin	13	8.55e-07	SQPSKKTYNA PSLVQRGKFARTTA
SRO15-2005	15	4.01e-06	IKQQKKAYVK PSMFQQGDFSKKTA

Supporting figure 10 – Alignment of the 27 leader peptides from the 30 known, functional lasso peptide biosynthetic gene clusters, wherein the motif was identified using the MEME algorithm; leader peptide of caulonodin V marked with a blue arrow, of rhodanodin with a red arrow

Supporting table 1 – Oligonucleotide primers

name	sequence
CK31_198x_A1_NdeI_FP	GGATTAC CAT ATG ACC CCG ATC CAA TCC AAG TTC TGC CTT CTG
CK31_198x_C_XhoI_RP	ATAT CTC GAG TCA TGA GAC CCT GAG CTT GGC TTC CCA GG
CK31_2238_A1_NdeI_FP	GGATTAC CAT ATG AAT ACG CTC AAG ACC CGG CTG ATC CG
CK31_2241_C_XhoI_RP	ATAT CTC GAG TTA GGC GCC GCG CGC TTC CCA G
CK31_198xB_SLIM_FPS	CTT TGG CTC GCC CAG GAC GTC CAT GC
CK31_198xB_rbs_SLIM_FPL	AG*AGGAGA*AATTAACC ATG CCC CTT TGG CTC GCC CAG GAC GTC CAT GC
CK31_198xA2_SLIM_RPS	TTA GGG CCA GTA GGT CTG CGA GCT CAT G
CK31_198xA2_rbs_SLIM_RPL	GGG CAT GGTTAATT*TCTCCT*CT TTA GGG CCA GTA GGT CTG CGA GCT CAT G
CK31_198xA2Del_SLIM_RPS	TCA GGC GAA GTA GTA TTG GCT GAC GAG ACC
CK31_198xA2Del_SLIM_RPL	GGG CAT GGTTAATT*TCTCCT*CT TCA GGC GAA GTA GTA TTG GCT GAC GAG ACC
CK31_198xA1Del_SLIM_FPS	ATG ACT CAA GTC TCT CCC TCG CCC CTG C
CK31_198xA1Del_SLIM_FPL	CTT TAA GAA GGA GAT ATA CAT ATG ACT CAA GTC TCT CCC TCG CCC CTG C
CK31_bothA1Del_SLIM_RPS	TTA AAC AAA ATT ATT TCT AGA GGG GAA TTG TTA TCC GCT CAC
CK31_bothA1Del_SLIM_RPL	ATG TAT ATC TCC TTC TTA AAG TTA AAC AAA ATT ATT TCT AGA GGG GAA TTG TTA TCC GCT CAC
CK31_22xxB_SLIM_FPS	CTA ACC TGG CGG CCC GGC GTT CAC
CK31_22xxB_SLIM_FPL	AG*AGGAGA*AATTAACC ATG ACG CTA ACC TGG CGG CCC GGC GTT CAC
CK31_22xxA2_SLIM_RPS	TCA GTC CCA ACG GCG AAC CAT GTT GGG C
CK31_22xxA2_SLIM_RPL	CGT CAT GGTTAATT*TCTCCT*CT TCA GTC CCA ACG GCG AAC CAT GTT GGG C
CK31_22xxA2Del_SLIM_RPS	TCA GGC CGG GTC GTA CCG CTT GAT CTG
CK31_22xxA2Del_SLIM_RPL	GGG CAT GGTTAATT*TCTCCT*CT TCA GGC CGG GTC GTA CCG CTT GAT CTG
CK31_22xxA1Del_SLIM_FPS	ATG ACG ACT CCC AAG TTT CGA CTG ATC CGT CTG
CK31_22xxA1Del_SLIM_FPL	CTT TAA GAA GGA GAT ATA CAT ATG ACG ACT CCC AAG TTT CGA CTG ATC CGT CTG
198xA12_SLIM_FPS	TGA AGA GGA GAA ATT AAC CAT GCC CCT TTG GCT C
1984A1_SLIM_RPS	GAG ACC TTC CTT GAT GGT GCC GAC ATC GAA G
1984A1_Y15A_FPL	GTC AGC CAA **GC**C TAC TTC GCC TGA AGA GGA GAA ATT AAC CAT GCC CCT TTG GCT C
1984A1_Y15A_RPL	GGC GAA GTA G**GC** TTG GCT GAC GAG ACC TTC CTT GAT GGT GCC GAC ATC GAA G
1984A1_Y16A_FPL	GTC AGC CAA TAC **GC**C TTC GCC TGA AGA GGA GAA ATT AAC CAT GCC CCT TTG GCT C
1984A1_Y16A_RPL	GGC GAA G**GC** GTA TTG GCT GAC GAG ACC TTC CTT GAT GGT GCC GAC ATC GAA G
1984A1_F17A_FPL	GTC AGC CAA TAC TAC **GC**C GCC TGA AGA GGA GAA ATT AAC CAT GCC CCT TTG GCT C
1984A1_F17A_RPL	GGC G**GC** GTA GTA TTG GCT GAC GAG ACC TTC CTT GAT GGT GCC GAC ATC GAA G
1984A1_Y15W_FPL	GTC AGC CAA T**GG** TAC TTC GCC TGA AGA GGA GAA ATT AAC CAT GCC CCT TTG GCT C
1984A1_Y15W_RPL	GGC GAA GTA **CC**A TTG GCT GAC GAG ACC TTC CTT GAT GGT GCC GAC ATC GAA G
1984A1_Y16W_FPL	GTC AGC CAA TAC T**GG** TTC GCC TGA AGA GGA GAA ATT AAC CAT GCC CCT TTG GCT C
1984A1_Y16W_RPL	GGC GAA **CC**A GTA TTG GCT GAC GAG ACC TTC CTT GAT GGT GCC GAC ATC GAA G
1984A1_F17W_FPL	GTC AGC CAA TAC TAC T**GG** GCC TGA AGA GGA GAA ATT AAC CAT GCC CCT TTG GCT C

name	sequence
1984A1_F17W_RPL	GGC **CC**A GTA GTA TTG GCT GAC GAG ACC TTC CTT GAT GGT GCC GAC ATC GAA G
198xA2_SLIM_RPS	CAT GGA CTC CCG CAG ACC GGA GTC G
198xA2_T15A_FPL	AGC TCG CAG **G**CC TAC TGG CCC TGA AGA GGA GAA ATT AAC CAT GCC CCT TTG GCT C
198xA2_T15A_RPL	GGG CCA GTA GG**C** CTG CGA GCT CAT GGA CTC CCG CAG ACC GGA GTC G
198xA2_Y16A_FPL	AGC TCG CAG ACC **GC**C TGG CCC TGA AGA GGA GAA ATT AAC CAT GCC CCT TTG GCT C
198xA2_Y16A_RPL	GGG CCA G**GC** GGT CTG CGA GCT CAT GGA CTC CCG CAG ACC GGA GTC G
198xA2_Y16F_FPL	AGC TCG CAG ACC T**T**C TGG CCC TGA AGA GGA GAA ATT AAC CAT GCC CCT TTG GCT C
198xA2_Y16F_RPL	GGG CCA G**A**A GGT CTG CGA GCT CAT GGA CTC CCG CAG ACC GGA GTC G
198xA2_W17A_FPL	AGC TCG CAG ACC TAC **GC**G CCC TGA AGA GGA GAA ATT AAC CAT GCC CCT TTG GCT C
198xA2_W17A_RPL	GGG C**GC** GTA GGT CTG CGA GCT CAT GGA CTC CCG CAG ACC GGA GTC G
198xA2_W17F_FPL	AGC TCG CAG ACC TAC T**TC** CCC TGA AGA GGA GAA ATT AAC CAT GCC CCT TTG GCT C
198xA2_W17F_RPL	GGG **GA**A GTA GGT CTG CGA GCT CAT GGA CTC CCG CAG ACC GGA GTC G
198xA2_W17Y_FPL	AGC TCG CAG ACC TAC T**AC** CCC TGA AGA GGA GAA ATT AAC CAT GCC CCT TTG GCT C
198xA2_W17Y_RPL	GGG **GT**A GTA GGT CTG CGA GCT CAT GGA CTC CCG CAG ACC GGA GTC G
CK31_223xA12_SLIM2_FPS	TGA AGA GGA GAA ATT AAC CAT GAC GCT AAC CTG GC
CK31_2238_SLIM2_RPS	CTG GTT GGT TTC CGG GAG AAG CAC CCC G
CK31_2238_K14A_FPL	ATC **GC**G CGG TAC GAC CCG GCC TGA AGA GGA GAA ATT AAC CAT GAC GCT AAC CTG GC
CK31_2238_K14A_RPL	GGC CGG GTC GTA CCG C**GC** GAT CTG GTT GGT TTC CGG GAG AAG CAC CCC G
CK31_2238_R15A_FPL	ATC AAG **GC**G TAC GAC CCG GCC TGA AGA GGA GAA ATT AAC CAT GAC GCT AAC CTG GC
CK31_2238_R15A_RPL	GGC CGG GTC GTA C**GC** CTT GAT CTG GTT GGT TTC CGG GAG AAG CAC CCC G
CK31_2238_Y16A_FPL	ATC AAG CGG **GC**C GAC CCG GCC TGA AGA GGA GAA ATT AAC CAT GAC GCT AAC CTG GC
CK31_2238_Y16A_RPL	GGC CGG GTC G**GC** CCG CTT GAT CTG GTT GGT TTC CGG GAG AAG CAC CCC G
CK31_2238_K14W_FPL	ATC **TG**G CGG TAC GAC CCG GCC TGA AGA GGA GAA ATT AAC CAT GAC GCT AAC CTG GC
CK31_2238_K14W_RPL	GGC CGG GTC GTA CCG C**CA** GAT CTG GTT GGT TTC CGG GAG AAG CAC CCC G
CK31_2238_R15W_FPL	ATC AAG **T**GG TAC GAC CCG GCC TGA AGA GGA GAA ATT AAC CAT GAC GCT AAC CTG GC
CK31_2238_R15W_RPL	GGC CGG GTC GTA CC**A** CTT GAT CTG GTT GGT TTC CGG GAG AAG CAC CCC G
CK31_2238_Y16W_FPL	ATC AAG CGG T**GG** GAC CCG GCC TGA AGA GGA GAA ATT AAC CAT GAC GCT AAC CTG GC
CK31_2238_Y16W_RPL	GGC CGG GTC **CC**A CCG CTT GAT CTG GTT GGT TTC CGG GAG AAG CAC CCC G
CK31_2239_SLIM_RPS	GGG CTC GGG GAA CAC GTC CCC GAT CC
CK31_2239_R14A_FPL	AAC ATG GTT **GC**C CGT TGG GAC TGA AGA GGA GAA ATT AAC CAT GAC GCT AAC CTG GC
CK31_2239_R14A_RPL	GTC CCA ACG G**GC** AAC CAT GTT GGG CTC GGG GAA CAC GTC CCC GAT CC
CK31_2239_R14W_FPL	AAC ATG GTT **T**G**G** CGT TGG GAC TGA AGA GGA GAA ATT AAC CAT GAC GCT AAC CTG GC

name	sequence
CK31_2239_R14W_RPL	GTC CCA ACG **C**C**A** AAC CAT GTT GGG CTC GGG GAA CAC GTC CCC GAT CC
CK31_2239_R15A_FPL	AAC ATG GTT CGC **GC**T TGG GAC TGA AGA GGA GAA ATT AAC CAT GAC GCT AAC CTG GC
CK31_2239_R15A_RPL	GTC CCA A**GC** GCG AAC CAT GTT GGG CTC GGG GAA CAC GTC CCC GAT CC
CK31_2239_R15K_FPL	AAC ATG GTT CGC **AAA** TGG GAC TGA AGA GGA GAA ATT AAC CAT GAC GCT AAC CTG GC
CK31_2239_R15K_RPL	GTC CCA **TTT** GCG AAC CAT GTT GGG CTC GGG GAA CAC GTC CCC GAT CC
CK31_2239_R15W_FPL	AAC ATG GTT CGC **T**G**G** TGG GAC TGA AGA GGA GAA ATT AAC CAT GAC GCT AAC CTG GC
CK31_2239_R15W_RPL	GTC CCA **C**C**A** GCG AAC CAT GTT GGG CTC GGG GAA CAC GTC CCC GAT CC
CK31_2239_W16A_FPL	AAC ATG GTT CGC CGT **GC**G GAC TGA AGA GGA GAA ATT AAC CAT GAC GCT AAC CTG GC
CK31_2239_W16A_RPL	GTC C**GC** ACG GCG AAC CAT GTT GGG CTC GGG GAA CAC GTC CCC GAT CC
CK31_2239_W16Y_FPL	AAC ATG GTT CGC CGT T**AC** GAC TGA AGA GGA GAA ATT AAC CAT GAC GCT AAC CTG GC
CK31_2239_W16Y_RPL	GTC **GT**A ACG GCG AAC CAT GTT GGG CTC GGG GAA CAC GTC CCC GAT CC
CK31_2239_W16F_FPL	AAC ATG GTT CGC CGT T**TC** GAC TGA AGA GGA GAA ATT AAC CAT GAC GCT AAC CTG GC
CK31_2239_W16F_RPL	GTC **GA**A ACG GCG AAC CAT GTT GGG CTC GGG GAA CAC GTC CCC GAT CC
198xA2_SLIM_RING_FPS	TCC GGT CTG CGG GAG TCC ATG AGC TC
198xA2_SLIM_RING_RPS	ATC GAG GGC GCG CCC GAC ACG
198xA2_S1G_FP	CTG ACC CGT **GG**G ATC GGC GAC TCC GGT CTG CGG GAG TCC ATG AGC TC
198xA2_S1G_RP	GTC GCC GAT C**CC** ACG GGT CAG ATC GAG GGC GCG CCC GAC ACG
198xA2 _S1A_FP	CTG ACC CGT **G**CG ATC GGC GAC TCC GGT CTG CGG GAG TCC ATG AGC TC
198xA2 _S1A_RP	GTC GCC GAT CG**C** ACG GGT CAG ATC GAG GGC GCG CCC GAC ACG
198xA2 _S1C_FP	CTG ACC CGT T**GC** ATC GGC GAC TCC GGT CTG CGG GAG TCC ATG AGC TC
198xA2 _S1C_RP	GTC GCC GAT **GC**A ACG GGT CAG ATC GAG GGC GCG CCC GAC ACG
198xA2 _S1F_FP	CTG ACC CGT T**TC** ATC GGC GAC TCC GGT CTG CGG GAG TCC ATG AGC TC
198xA2 _S1F_RP	GTC GCC GAT **GA**A ACG GGT CAG ATC GAG GGC GCG CCC GAC ACG
198xA2_S1V_FP	CTG ACC CGT **GT**G ATC GGC GAC TCC GGT CTG CGG GAG TCC ATG AGC TC
198xA2_S1V_RP	GTC GCC GAT C**AC** ACG GGT CAG ATC GAG GGC GCG CCC GAC ACG
198xA2_S1T_FP	CTG ACC CGT **A**CG ATC GGC GAC TCC GGT CTG CGG GAG TCC ATG AGC TC
198xA2_S1T_RP	GTC GCC GAT CG**T** ACG GGT CAG ATC GAG GGC GCG CCC GAC ACG
CK31_2238_SLIM1_FPS	GTG CTT CTC CCG GAA ACC AAC CAG ATC AAG C
CK31_2238_SLIM1_RPS	GCG CTT GGC CGA GCC GAA ACG G
CK31_2238_A1G_FPL	CTG ACC CGC G**G**C GGC ACC GGG GTG CTT CTC CCG GAA ACC AAC CAG ATC AAG C
CK31_2238_A1G_RPL	CCC GGT GCC G**C**C GCG GGT CAG GCG CTT GGC CGA GCC GAA ACG G
CK31_2238_A1S_FPL	CTG ACC CGC **T**CC GGC ACC GGG GTG CTT CTC CCG GAA ACC AAC CAG ATC AAG C

name	sequence
CK31_2238_A1S_RPL	CCC GGT GCC GG**A** GCG GGT CAG GCG CTT GGC CGA GCC GAA ACG G
CK31_2238_A1C_FPL	CTG ACC CGC **TG**C GGC ACC GGG GTG CTT CTC CCG GAA ACC AAC CAG ATC AAG C
CK31_2238_A1C_RPL	CCC GGT GCC G**CA** GCG GGT CAG GCG CTT GGC CGA GCC GAA ACG G
CK31_2238_A1F_FPL	CTG ACC CGC **TT**C GGC ACC GGG GTG CTT CTC CCG GAA ACC AAC CAG ATC AAG C
CK31_2238_A1F_RPL	CCC GGT GCC G**AA** GCG GGT CAG GCG CTT GGC CGA GCC GAA ACG G
CK31_2238_A1V_FPL	CTG ACC CGC G**T**C GGC ACC GGG GTG CTT CTC CCG GAA ACC AAC CAG ATC AAG C
CK31_2238_A1V_RPL	CCC GGT GCC G**A**C GCG GGT CAG GCG CTT GGC CGA GCC GAA ACG G
CK31_2238_A1T_FPL	CTG ACC CGC **A**CC GGC ACC GGG GTG CTT CTC CCG GAA ACC AAC CAG ATC AAG C
CK31_2238_A1T_RPL	CCC GGT GCC GG**T** GCG GGT CAG GCG CTT GGC CGA GCC GAA ACG G
198xLeaderDel1_RtH_RP	CAT ATG TAT ATC TCC TTC TTA AAG TTA AAC AAA ATT ATT TCT AGA GGG GAA TTG
198xLeaderDel1_RtH_FP	CTG CGG CTC ATC CGT GTC GGG CG
198xLeaderDel12_RtH_FP	GGG CGC GCC CTC GAT CTG ACC CG
198xLeaderDel123_RtH_FP	ACC CGT TCG ATC GGC GAC TCC GGT CTG
198xLeaderDel2FPS	CTC GAT CTG ACC CGT TCG ATC GGC GAC
198xLeaderDel2RPS	AGA GAC TTG AGT CAT ATG TAT ATC TCC TTC TTA AAG TTA AAC AAA ATT ATT TC
198xLeaderDel2FPL	CCC TCG CCC GGG CGC GCC CTC GAT CTG ACC CGT TCG ATC GGC GAC
198xLeaderDel2RPL	GGC GCG CCC GGG CGA GGG AGA GAC TTG AGT CAT ATG TAT ATC TCC TTC TTA AAG TTA AAC AAA ATT ATT TC
198xLeaderDel3FPS	ATC GGC GAC TCC GGT CTG CGG GAG TCC
198xLeaderDel3RPS	GAG CCG CAG GGG CGA GGG AGA GAC TTG
198xLeaderDel3FPL	ATC CGT GTC ACC CGT TCG ATC GGC GAC TCC GGT CTG CGG GAG TCC
198xLeaderDel3RPL	CGA ACG GGT GAC ACG GAT GAG CCG CAG GGG CGA GGG AGA GAC TTG
198xLeaderBlock1 SLIM FPS	GAT CTG ACC CGT TCG ATC GGC GAC TCC
198xLeaderBlock1 SLIM RPS	CCG CAG GGG CGA GGG AGA GAC TTG
198xLeaderLIVAtoRRRR FPL	**CGC CGT** CGT **CGC** GGG CGC **CGT** CTC GAT CTG ACC CGT TCG ATC GGC GAC TCC
198xLeaderLIVAtoRRRR RPL	GAG **ACG** GCG CCC **GCG** ACG **ACG GCG** CCG CAG GGG CGA GGG AGA GAC TTG
198xLeaderLIVAtoEEEE FPL	**GAA GAG** CGT **GAA** GGG CGC **GAA** CTC GAT CTG ACC CGT TCG ATC GGC GAC TCC
198xLeaderLIVAtoEEEE RPL	GAG **TTC** GCG CCC **TTC** ACG **CTC TTC** CCG CAG GGG CGA GGG AGA GAC TTG
198x_LI-11-12toRR FPL	**CGC CGT** CGT GTC GGG CGC GCC CTC GAT CTG ACC CGT TCG ATC GGC GAC TCC
198x_LI-11-12toRR RPL	GAG GGC GCG CCC GAC ACG **ACG GCG** CCG CAG GGG CGA GGG AGA GAC TTG
198x_LI-11-12toEE FPL	**GAA GAG** CGT GTC GGG CGC GCC CTC GAT CTG ACC CGT TCG ATC GGC GAC TCC
198x_LI-11-12toEE RPL	GAG GGC GCG CCC GAC ACG **CTC TTC** CCG CAG GGG CGA GGG AGA GAC TTG
198x_LI-11-12toQQ FPL	**CAA CAG** CGT GTC GGG CGC GCC CTC GAT CTG ACC CGT TCG ATC GGC GAC TCC
198x_LI-11-12toQQ RPL	GAG GGC GCG CCC GAC ACG **CTG TTG** CCG CAG GGG CGA GGG AGA GAC TTG
198x_VA-9-6toRR FPL	CTC ATC CGT **CGC** GGG CGC **CGT** CTC GAT CTG ACC CGT TCG ATC GGC GAC TCC

name	sequence
198x_VA-9-6toRR RPL	GAG **ACG** GCG CCC **GCG** ACG GAT GAG CCG CAG GGG CGA GGG AGA GAC TTG
198x_VA-9-6toEE FPL	CTC ATC CGT **GAA** GGG CGC **GAA** CTC GAT CTG ACC CGT TCG ATC GGC GAC TCC
198x_VA-9-6toEE RPL	GAG **TTC** GCG CCC **TTC** ACG GAT GAG CCG CAG GGG CGA GGG AGA GAC TTG
198x_VA-9-6toQQ FPL	CTC ATC CGT **CAG** GGG CGC **CAG** CTC GAT CTG ACC CGT TCG ATC GGC GAC TCC
198x_VA-9-6toQQ RPL	GAG **CTG** GCG CCC **CTG** ACG GAT GAG CCG CAG GGG CGA GGG AGA GAC TTG
198x_V-9R FPL	CTC ATC CGT **CGC** GGG CGC GCC CTC GAT CTG ACC CGT TCG ATC GGC GAC TCC
198x_V-9R RPL	GAG GGC GCG CCC **GCG** ACG GAT GAG CCG CAG GGG CGA GGG AGA GAC TTG
198x_V-9E FPL	CTC ATC CGT **GAA** GGG CGC GCC CTC GAT CTG ACC CGT TCG ATC GGC GAC TCC
198x_V-9E RPL	GAG GGC GCG CCC **TTC** ACG GAT GAG CCG CAG GGG CGA GGG AGA GAC TTG
198x_A-6R FPL	CTC ATC CGT GTC GGG CGC **CGT** CTC GAT CTG ACC CGT TCG ATC GGC GAC TCC
198x_A-6R RPL	GAG **ACG** GCG CCC GAC ACG GAT GAG CCG CAG GGG CGA GGG AGA GAC TTG
198x_A-6E FPL	CTC ATC CGT GTC GGG CGC **GAA** CTC GAT CTG ACC CGT TCG ATC GGC GAC TCC
198x_A-6E RPL	GAG **TTC** GCG CCC GAC ACG GAT GAG CCG CAG GGG CGA GGG AGA GAC TTG
198x_G-8F FPL	CTC ATC CGT GTC **TTC** CGC GCC CTC GAT CTG ACC CGT TCG ATC GGC GAC TCC
198x_G-8F RPL	GAG GGC GCG **GAA** GAC ACG GAT GAG CCG CAG GGG CGA GGG AGA GAC TTG
198x_G-8R FPL	CTC ATC CGT GTC **CGT** CGC GCC CTC GAT CTG ACC CGT TCG ATC GGC GAC TCC
198x_G-8R RPL	GAG GGC GCG **ACG** GAC ACG GAT GAG CCG CAG GGG CGA GGG AGA GAC TTG
198x_G-8E FPL	CTC ATC CGT GTC **GAA** CGC GCC CTC GAT CTG ACC CGT TCG ATC GGC GAC TCC
198x_G-8E RPL	GAG GGC GCG **TTC** GAC ACG GAT GAG CCG CAG GGG CGA GGG AGA GAC TTG
198x_G-8A FPL	CTC ATC CGT GTC **GCG** CGC GCC CTC GAT CTG ACC CGT TCG ATC GGC GAC TCC
198x_G-8A RPL	GAG GGC GCG **CGC** GAC ACG GAT GAG CCG CAG GGG CGA GGG AGA GAC TTG
198x_G-8V FPL	CTC ATC CGT GTC **GTG** CGC GCC CTC GAT CTG ACC CGT TCG ATC GGC GAC TC
198x_G-8V RPL	GAG GGC GCG **CAC** GAC ACG GAT GAG CCG CAG GGG CGA GGG AGA GAC TTG
198x_G-8P FPL	CTC ATC CGT GTC **CCG** CGC GCC CTC GAT CTG ACC CGT TCG ATC GGC GAC TCC
198x_G-8P RPL	GAG GGC GCG **CGG** GAC ACG GAT GAG CCG CAG GGG CGA GGG AGA GAC TTG
198x_R-10A FPL	CTC ATC **GCT** GTC GGG CGC GCC CTC GAT CTG ACC CGT TCG ATC GGC GAC TCC
198x_R-10A RPL	GAG GGC GCG CCC GAC **AGC** GAT GAG CCG CAG GGG CGA GGG AGA GAC TTG
198x_R-10F FPL	CTC ATC **TTT** GTC GGG CGC GCC CTC GAT CTG ACC CGT TCG ATC GGC GAC TCC
198x_R-10F RPL	GAG GGC GCG CCC GAC **AAA** GAT GAG CCG CAG GGG CGA GGG AGA GAC TTG
198x_R-10E FPL	CTC ATC **GAA** GTC GGG CGC GCC CTC GAT CTG ACC CGT TCG ATC GGC GAC TCC

name	sequence
198x_R-10E RPL	GAG GGC GCG CCC GAC **TTC** GAT GAG CCG CAG GGG CGA GGG AGA GAC TTG
198xLeaderBlock 2SLIM FPS	TCC GGT CTG CGG GAG TCC ATG AGC TC
198xLeaderBlock 2SLIM RPS	GGC GCG CCC GAC ACG GAT GAG C
198xR-1A_ FPL	CTC GAT CTG ACC **GCC** TCG ATC GGC GAC TCC GGT CTG CGG GAG TCC ATG AGC TC
198xR-1A_ RPL	GTC GCC GAT CGA **GGC** GGT CAG ATC GAG GGC GCG CCC GAC ACG GAT GAG C
198xR-1Q_ FPL	CTC GAT CTG ACC **CAG** TCG ATC GGC GAC TCC GGT CTG CGG GAG TCC ATG AGC TC
198xR-1Q_ RPL	GTC GCC GAT CGA **CTG** GGT CAG ATC GAG GGC GCG CCC GAC ACG GAT GAG C
198xR-1N_ FPL	CTC GAT CTG ACC **AAC** TCG ATC GGC GAC TCC GGT CTG CGG GAG TCC ATG AGC TC
198xR-1N_ RPL	GTC GCC GAT CGA **GTT** GGT CAG ATC GAG GGC GCG CCC GAC ACG GAT GAG C
198xR-1K_ FPL	CTC GAT CTG ACC **AAG** TCG ATC GGC GAC TCC GGT CTG CGG GAG TCC ATG AGC TC
198xR-1K_ RPL	GTC GCC GAT CGA **CTT** GGT CAG ATC GAG GGC GCG CCC GAC ACG GAT GAG C
198xR-1H_ FPL	CTC GAT CTG ACC **CAC** TCG ATC GGC GAC TCC GGT CTG CGG GAG TCC ATG AGC TC
198xR-1H_ RPL	GTC GCC GAT CGA **GTG** GGT CAG ATC GAG GGC GCG CCC GAC ACG GAT GAG C
198xR-1F_ FPL	CTC GAT CTG ACC **TTC** TCG ATC GGC GAC TCC GGT CTG CGG GAG TCC ATG AGC TC
198xR-1F_ RPL	GTC GCC GAT CGA **GAA** GGT CAG ATC GAG GGC GCG CCC GAC ACG GAT GAG C
198xR-1E_ FPL	CTC GAT CTG ACC **GAA** TCG ATC GGC GAC TCC GGT CTG CGG GAG TCC ATG AGC TC
198xR-1E_ RPL	GTC GCC GAT CGA **TTC** GGT CAG ATC GAG GGC GCG CCC GAC ACG GAT GAG C
198xT-2A_ FPL	CTC GAT CTG **GCC** CGT TCG ATC GGC GAC TCC GGT CTG CGG GAG TCC ATG AGC TC
198xT-2A_ RPL	GTC GCC GAT CGA ACG **GGC** CAG ATC GAG GGC GCG CCC GAC ACG GAT GAG C
FP_Rhothi_Leader_SLIM	GGA GTA TTG CCC ATC GGC AAT GAG TTC ATG GGC CAC G
RP_Rhothi_Leader_SLIM	GAG CTC GAT GAC ATC GTC CTG CGC GTT GCT G
FPTail_Rhothi_S-8G_SLIM	**GGC** GTC GCC AGT GTC GAG ACC AAA GGA GTA TTG CCC ATC GGC AAT GAG TTC ATG GGC CAC G
RPTail_Rhothi_S-8G_SLIM	TTT GGT CTC GAC ACT GGC GAC **GCC** GAG CTC GAT GAC ATC GTC CTG CGC GTT GCT G

Introduced restriction enzyme sites are underlined; introduced ribosomal binding sites are shown in italics; introduced mutations are shown in bold

Supporting table 2 – M9 Vitamin Mix

component	amount
biotin	0.2 g
choline chloride	1.0 g
disodium adenosine 5’-triphosphate	0.3 g
folic acid	1.0 g
myo-inositol	2.0 g
nicotinamide	1.0 g
panthothenic acid	1.0 g
pyridoxal hydrochloride	1.0 g
riboflavin	0.1 g
thiamine	1.0 g
H_2O	ad 300 mL

10 M NaOH were added to the resulting suspension until everything was dissolved. The solution was then sterile filtered and stored at 4°C

Supporting Table 3 – Tail sequences of the caulonodins with possible plug residues highlighted

peptide	14	15	16	17	18	19
caulonodin IV	Q	**Y**	**Y**	**F**	A	-
caulonodin V	Q	**T**	**Y**	**W**	P	-
caulonodin VI	**K**	**R**	**Y**	D	P	A
caulonodin VII	**R**	**R**	**W**	D	-	-

Positions picked for mutation in bold black, most likely plugs in red

Supporting Table 4 – Production amounts of all constructed caulonodin and rhodanodin mutants; mutant numbering according to figure 1a, positive numbers for mutations in the core peptide, negative numbers for mutations in the leader peptide.

lasso peptide mutant	relative production to corresponding wild type in %	heat stable
caulonodin IV WT	100 ± 15	no
Y15A	49 ± 2	no
Y15W	167 ± 12	no
Y16A	14.3 ± 0.7	no
Y16W	84 ± 6	no
F17A	not detected	-
F17W	132 ± 9	yes
caulonodin V WT	100 ± 23	no
S1G	4.0 ± 0.3	-
S1A	132.2 ± 1.6	-
S1C	detected	-
S1V	detected	-
S1T	detected	-
S1F	detected	-
T15A	2.7 ± 0.1	-
Y16A	detected	-
Y16F	detected	-
W17A	detected	-
W17F	detected	-
W17Y	8.5 ± 0.6	-
Del1	7.8 ± 0.5	-
Del2	detected	-
Del3	detected	-
Del12	not detected	-
Del123	not detected	-
L-12R_I-11R_V-9R_A-6R	not detected	-
L-12E_I-11E_V-9E_A-6E	not detected	-
L-12R_I-11R	0.3 ± 0.1	-
L-12E_I-11E	0.4 ± 0.1	-
L-12Q_I-11Q	7.6 ± 0.1	-
V-9R_A-6R	not detected	-
V-9E_A-6E	not detected	-
V-9Q_A-6Q	not detected	-
R-10A	97.8 ± 1.8	-
R-10F	71.2 ± 1.5	-
R-10E	63.0 ± 1.4	-
V-9E	2.0 ± 0.1	-
V-9R	3.6 ± 0.2	-
G-8A	detected	-

lasso peptide mutant	relative production to corresponding wild type in %	heat stable
G-8F	6.0 ± 0.5	-
G-8E	4.7 ± 0.1	-
G-8V	0.5 ± 0.1	-
G-8R	not detected	-
G-8P	0.3 ± 0.1	-
A-6E	detected	-
A-6R	detected	-
T-2A	detected	-
R-1A	80 ± 15	-
R-1F	1.0 ± 0.1	-
R-1E	33 ± 14	-
R-1Q	84 ± 3	-
R-1N	62 ± 12	-
R-1K	90 ± 8	-
R-1H	15.7 ± 1.0	-
caulonodin VI WT	100 ± 7	no
A1G	7.1 ± 0.5	-
A1S	122 ± 30	-
A1C	detected	-
A1V	detected	-
A1T	1.3 ± 0.1	-
A1F	detected	-
K14A	38 ± 6	no
K14W	22.9 ± 1.4	no
R15A	6.1 ± 0.5	no
R15W	2.0 ± 0.1	yes
Y16A	detected	-
Y16W	99 ± 14	no
caulonodin VII WT	100 ± 29	no
R14A	29.7 ± 0.6	no
R15A	0.8 ± 0.1	no
R15K	123 ± 5	no
R15W	detected	yes
W16A	not detected	-
W16F	detected	-
W16Y	73 ± 12	no
rhodanodin WT	100	-
S-8G	1460	-

Supporting table 5 – Chemical shift Assignment of ^{1}H signals (ppm) of caulonodin V in H_2O/D_2O (9:1) at 298 K

aa	NH	αH	βH	others
Ser1	7.075	4.560	4.190; 3.995	/
Ile2	9.084	4.891	2.509	γCH_3: 1.159
				γCH_2: 1.478; 1.342
				δCH_3: 1.151
Gly3	8.702	4.528; 3.664	/	/
Asp4	8.973	5.155	3.082	/
Ser5	7.600	3.240	3.821	/
Gly6	8.185	4.651; 3.701	/	/
Leu7	9.406	4.841	2.089	γCH: 2.044
				δCH_3: 1.244; 1.172
Arg8	8.872	4.007	1.863; 1.737	γCH_2: 1.700; 1.513
				δCH_2: 3.170; 3.132
				εNH: 7.140
Glu9	7.989	3.386	1.485; 1.327	γCH_2: 1.225
Ser10	7.854	4.691	4.200; 4.008	/
Met11	8.900	4.574	2.304	γCH_2: 2.894; 2.763
				εCH_3: 2.290
Ser12	8.305	4.730	4.078; 3.969	/
Ser13	7.671	4.653	4.013; 3.871	/
Gln14	8.921	4.488	2.039	γCH_2: 2.494
				εNH_2: 7.649; 6.992
Thr15	7.468	4.643	4.204	γCH_3: 1.221
Tyr16	8.058	5.136	3.033; 2.165	2. 6H: 6.787
				3. 5H: 6.787
Trp17	9.350	5.488	3.731; 3.293	2H: 7.176; 4H: 7.941
				5H: 7.278; 6H: 7.434
				7H: 7.779; NH: 10.401
Pro18	/	4.774	2.516; 2.355	γCH2: 2.267; 2.468
				δCH_2: 4.483; 4.351

Supporting table 6 – Structural statistics for the family of 15 structures selected to represent the structure of caulonodin V in H_2O/D_2O (9:1)

Restraining Constraints	Constraints Violations
Total: 166	Distance violations, >0.5 Å: 0
distance, i=j: 79	RMS deviations: 0.02 Å
distance, |i-j|=1: 35	Dihedral violations, > 5°: 0
distance, |i-j|>1: 22	RMS deviation: 1.8°
dihedral: 22	Average pairwise RMS deviation (Ile^2-Trp^{17})
hydrogen bond: 0	Backbone atoms: 0.06 Å
Constraints/residue: 9.2	All heavy atoms: 0.45 Å

Supporting table 7 – Chemical shift Assignment of ^{1}H signals (ppm) of caulonodin VI in H_2O/D_2O (9:1) at 288 K

aa	NH	αH	βH	others
Ala1	7.654	4.426	1.617	/
Gly2	8.608	4.576; 4.602		
Thr3	7.695	4.623	4.437	γCH3: 1.358
Gly4	7.700	3.844; 4.524		
Val5	7.960	4.829	2.516	γCH3: 0.964; 1.008
Leu6	8.027	4.289	amb.*	γCH2: amb.
				δCH3: 1.038; 1.078
Leu7	8.962	5.131	amb.	γCH2 amb.
				δCH3: 0.911; 1.096
Pro8	/	4.776	amb.	Amb.
Glu9	8.790	4.586	2.242	γCH2: 2.275; 2.424
Thr10	8.659	4.293	4.146	γCH3: 1.444
Asn11	amb.	amb.	amb.	amb.
Gln12	8.252	4.384	2.340; 2.547	γCH2: 2.505; 2.575
				εNH2: amb.
Ile13	8.412	4.494	amb.	γCH2: amb.
				γCH3: 1.120; δCH3: 1.056
Lys14	7.989	4.756	amb.	γCH2: amb.; δCH2: amb.
				εCH2: 3.162; $\varepsilon NH3^{+}$: 7.825
Arg15	8.521	5.593	amb.	γCH2: amb.
				δCH2: 3.077; 3.479
				NH: 7.143
Tyr16	9.772	4.727	3.135; 3.183	2, 6H: 7.027
				3, 5H: 6.825
Asp17	8.959	5.363	amb.	
Pro18	/	4.693	amb.	amb.
Ala19	8.825	4.532	1.644	

*amb. = ambiguous

Rational Improvement of the Affinity and Selectivity of Integrin Binding of Grafted Lasso Peptides

Julian D. Hegemann,[a] Mariarosaria De Simone,[b,†] Marcel Zimmermann,[a] Thomas A. Knappe,[a,††] Xiulan Xie,[a] Francesco Saverio Di Leva,[c] Ettore Novellino,[c] Luciana Marinelli,[c] Stefan Zahler,[d] Horst Kessler[b] and Mohamed A. Marahiel[a,*]

[a]*Department of Chemistry, Biochemistry, Philipps-Universiät Marburg, Hans-Meerwein-Strasse 4 and LOEWE-Center for Synthetic Microbiology, 35032 Marburg, Germany*

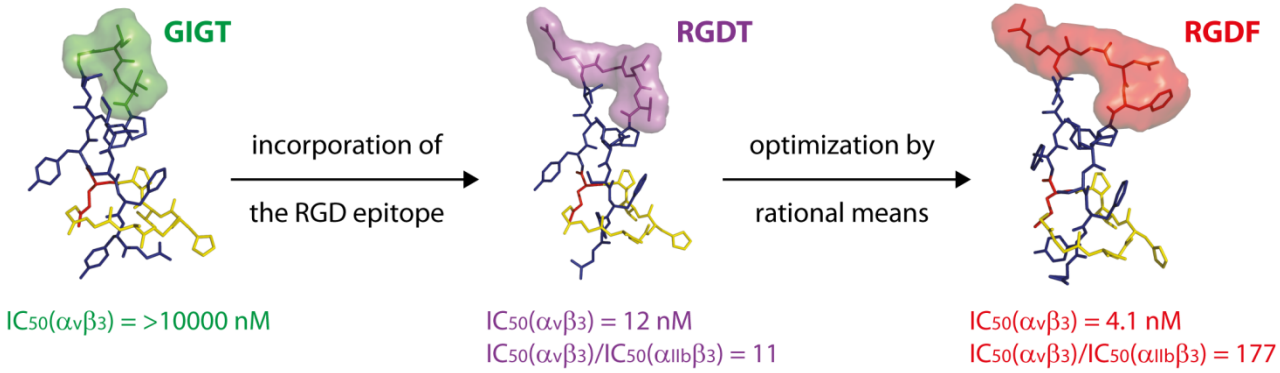
[b]*Institute for Advanced Study, Center of Integrated Protein Science (CIPSM), Department of Chemistry, Lichtenbergstrasse 4, Technische Universität München, 85747 Garching, Germany and Chemistry Department, Faculty of Science, King Abdulaziz University, P.O. Box 80203, Jeddah 21589, Saudi Arabia*

[c]*Department of Pharmacy, Universita` di Napoli "Federico II", Via D. Montesano 49, 80131 Napoli, Italy*

[d]*Department of Pharmacy, Ludwig-Maximilians-University, Butenandtstrasse 5-13, 81377 Munich, Germany*

† Dr. Mariarosaria De Simone, Istituto Di Fisiologia Clinica (IFC-CNR), Via Moruzzi 1. Pisa 56124, Italy.

†† Dr. Thomas Knappe, Ascendis Pharma GmbH, Im Neuenheimer Feld 584, 69120 Heidelberg, Germany.



ABSTRACT: Integrins moderate diverse important functions in the human body and are promising targets in cancer therapy. Hence, the selective inhibition of specific integrins is of great medicinal interest. Here, we report the optimization of a grafted lasso peptide, yielding MccJ25(RGDF), which is a highly potent and selective inhibitor of the αvβ3 integrin. Furthermore, its elucidated structure was employed in a molecular dynamics approach, revealing information about the integrin binding mode and selectivity profile of MccJ25(RGDF).

Supporting Information

Table of Contents

Experimental Procedures

Generation, Production and Isolation of MccJ25 Variants. Oligonucleotide primers were purchased from Sigma Aldrich. *E. coli* TOP10 was used for the generation of the mutants, while *E. coli* BL21(DE3) was used for the heterologous expression. Both strains were obtained from Invitrogen. Phusion polymerase was purchased from New England Biolabs. For confirming the correct incorporation of the desired mutations, DNA dideoxy sequencing was performed by GATC Biotech. Variants of the plasmid pTUC202[1] were generated by employing site-directed ligase-independent mutagenesis (SLIM)[2, 3] using the oligonucleotide primers shown in Supplementary Table S1. Initial screens of the production of the new MccJ25 variants were done in 100 mL M63 minimal medium (13.6 g/L KH_2PO_4, 2 g/L $(NH_4)_2SO_4$, 1 g/L bacto casamino acids, 1 mL/L $MgSO_4$ solution (2 M), pH = 7.0; after autoclaving 5 mL/L glucose solution (40% w/v), 0.2 mL/L biotin solution (10 mg/L) and 0.2 mL/L thiamine solution (10 mg/mL) were added) containing 17 μg/mL chloramphenicol that was inoculated to an OD_{600} of 0.01 with LB overnight cultures containing the same antibiotic. After fermentation for 1 day at 37 °C, the cells were harvested by centrifugation. Supernatants were collected and applied to solid phase extraction using XAD16 resin (Sigma Aldrich, ~4 g/L culture supernatant). After the addition of the XAD16 resin, the supernatant was stirred for 1 h at room temperature. Subsequently, the resin was removed by filtration, washed with water and eluted with pure MeOH (100 mL/L of the initial culture volume). Solvent was evaporated at 40 °C under reduced pressure and the dried extracts were each resuspended in 1 mL of 50% MeOH. These extracts were cleared by centrifugation and analyzed with an LTQ-FT ultra instrument (Thermo Fisher Scientific) connected to a microbore 1100 HPLC system (Agilent) using 100 μL of extract for each measurement. For detection the UV absorption at 215 nm was recorded. Separation was achieved using a CC 125/2 Nucleosil 300-8 C18 column (Macherey-Nagel) applying the following gradient of water/0.1% trifluoroacetic acid (solvent A) and MeCN/0.1% trifluoroacetic acid (solvent B) at a column temperature of 40 °C and a flow rate of 0.2 mL/min: holding 2% B for 2 min, followed by a linear increase from 2% to 30% B in 18 min, a subsequent linear increase from 30% to 95% B in 15 min and holding 95% B for additional 2 min. In both cases, the predicted masses corresponding to the full length lasso peptides were detected.

Subsequent large scale fermentation and isolation of the novel MccJ25 variants was then performed under the tested conditions. After XAD16 extraction, MccJ25 variants were purified by two rounds of preparative HPLC (microbore 1100 HPLC system (Agilent) with a VP 250/21 Nucleodur C18 Htec 5 μm column (Macherey-Nagel) at room temperature and a flow rate of 18 mL/min).

For the first round, the following gradient of water/0.05% formic acid (solvent A) and

MeOH/0.04% formic acid (solvent B) was used: Linear increase from 40% to 55% B in 30 min, followed by a linear increase from 55% to 95% B in 2 min and holding 95% B for an additional 3 min. The observed retention times were 25.0 min for MccJ25(FRGD) and 27.0 min for MccJ25(RGDF). The pooled fractions of each variant were evaporated at 40 °C under reduced pressure and reapplied to a second round of purification using the following gradient of water/0.1% trifluoroacetic acid (solvent C) and MeCN/0.1% trifluoroacetic acid (solvent D): Linear increase from 20% to 40% D in 30 min, followed by a linear increase from 40% to 95% D in 2 min and holding 95% D for additional 3 min. The observed retention times were 26.5 min for MccJ25(FRGD) and 25.5 min for MccJ25(RGDF). Pure MccJ25(FRGD) was obtained with a yield of 0.2 mg/L culture, while pure MccJ25(RGDF) could be isolated with a yield of 4.6 mg/L culture. MccJ25(RGD) was produced, isolated and purified as previously described.[4] The purity of the isolated compounds was assessed by using a microbore 1100 HPLC system (Agilent) coupled with a 1100 series MSD (Hewlett-Packard). For this, a sample of each compound was separated over an EC 125/2 Nucleodur 100 3 C18ec column and checked for impurities. In all instances, the LC MS analysis showed the compounds to have a purity that was equal to or even exceeding 95%.

Assessment of the I*n Vitro* Integrin Inhibition by ELISA. The inhibiting activity and selectivity of the tested antagonists were determined by using ELISAs based on previously reported methods with some optimizations.[5, 6]

Human integrins αvβ3 and αvβ5 were purchased from Millipore, αIIbβ3 from Enzyme Research Laboratories, α5β1 and αvβ3 from R&D Systems. Vitronectin was purchased from Millipore, fibronectin from Sigma Aldrich, fibrinogen from Calbiochem and LAP protein from R&D Systems. For the integrins αvβ3 and αvβ5 the binding was visualized using a mouse anti-human integrin α-V monoclonal antibody for the αv subunit (MAB 1978 purchased from Chemicon) while for α5β1 and αIIbβ3 antibodies from BD Biosciences were used (mouse anti-human CD49e for α5β1, mouse anti-human CD41b for αIIbβ3). In all cases, anti-mouse IgG-peroxidase from Sigma Aldrich was used as secondary antibody, containing a peroxidase conjugate that was employed for the visualization and quantification. Peroxidase development was performed by using the substrate solution 3,3,5,5’-tetramethylethylenediamine (TMB) from Seramun Diagnostic GmbH and by adding 3 M H_2SO_4 to stop the reaction.

The absorbance (450, 492 nm) was recorded with a POLARstar Galaxy plate reader (BMG Labtechnologies). Every concentration was analyzed in duplicate and the resulting inhibition curves were analyzed using OriginPro 7.5G software. The turning point of these curves describes the IC_{50} value. Each plate contained either cilengitide or tirofiban as reference compounds.[7, 8]

Blocking and binding steps were performed with TS buffer (20 mM Tris-HCl pH 7.5, 150 mM NaCl,

1 mM $CaCl_2$, 1 mM $MgCl_2$, and 1 mM $MnCl_2$) containing 1% BSA (TSB buffer). After the incubation time, washing steps were done with PBST buffer (10 mM Na_2HPO_4, pH 7.5, 150 mM NaCl, and 0.01% Tween20).

For the αvβ3 assays, a flat-bottom 96-well ELISA plate (from Brand) was coated overnight with 100 µL of vitronectin (2 µg/mL) in carbonate buffer (15 mM Na_2CO_3, 35 mM $NaHCO_3$, pH 9.6) at a temperature of 4 °C. After removing the solutions from the plate, the wells were blocked for 1 h at room temperature with 150 µL TSB buffer per well. The plate was subsequently washed three times with 200 µL PBST buffer per well. Afterward, 4.0 µg/mL of the soluble integrin αvβ3 and a serial dilution of the different compounds and a cilengitide control were incubated in the coated wells for 2 h at room temperature. After washing three times, the plate was treated with each 100 µL of the primary antibody (MAB1978, diluted 1:500 in TSB buffer, 1.0 µg/mL) and 100 µL of the secondary antibody (anti-mouse IgG-peroxidase, diluted 1:385 in TSB buffer, 2.0 µg/mL) per well for 1 h at room temperature. After this treatment, the plate was washed three times and the binding was then visualized using TMB. For this substrate, the oxidation reaction was performed for only 5 min and the absorbance was measured at 450 nm. IC_{50} determination was done as explained above.

For the αvβ5 assays, an appropriate 96-well ELISA plate was coated overnight with 100 µL of vitronectin (5 µg/mL) in TS buffer at a temperature of 4 °C. After removing the solutions from the plate, the wells were blocked for 1 h at room temperature with 150 µL TSB buffer per well. Plates were then washed three times with 200 µL PBST buffer per well. Subsequently, 3.0 µg/mL of the soluble integrin αvβ5 and a serial dilution of the different compounds and a cilengitide control were incubated in the coated wells for 1 h at room temperature. After washing three times, the plate was treated with each 100 µL of the primary antibody (MAB1978, diluted 1:500 in TSB buffer, 1.0 µg/mL) and 100 µL of the secondary antibody (anti-mouse IgG-peroxidase, diluted 1:770 in TSB buffer, 1.0 µg/mL) per well for 1 h at room temperature. Then, the plate was treated with TMB and the absorbance was measured as previously described.

For the α5β1 assays, the flat-bottom 96-well ELISA plate was coated overnight at 4 °C with 100 µL of fibronectin (0.5 µg/mL) in carbonate buffer (15 mM Na_2CO_3, 35 mM $NaHCO_3$, pH 9.6). The plates were subsequently blocked for 1 h at room temperature with 150 µL TSB buffer per well. Afterward, 1.0 µg/mL of the soluble integrin α5β1 and a serial dilution of the different compounds and a cilengitide control were incubated in the coated wells for 1 h at room temperature. After washing three times, the plate was treated with each 100 µL of the primary antibody (CD49e, diluted 1:500 in TSB buffer, 1.0 µg/mL) and 100 µL of the secondary antibody (anti-mouse IgG-peroxidase, diluted 1:385 in TSB buffer, 2.0 µg/mL) per well for 1 h at room temperature. After this treatment, the plate was washed three times and the binding was visualized with TMB as described above. Then, the absorbance was measured at 450 nm and the IC_{50} values were

determined as mentioned before.

For the αIIbβ3 assays, a flat-bottom 96-well ELISA plate was coated overnight at 4 °C with 100 µL of fibronogen (10.0 µg/mL) in carbonate buffer (15 mM Na_2CO_3, 35 mM $NaHCO_3$, pH 9.6). After the plates were blocked for 1 h at room temperature with 150 µL TSB buffer per well, 2.5 µg/mL of the soluble integrin αIIbβ3 and a serial dilution of the different compounds and a cilengitide as well as a tirofiban control were incubated in the coated wells for 1 h at room temperature. The plate was then washed three times and subsequently treated with each 100 µL of the primary antibody (CD41b, diluted 1:250 in TSB buffer, 2.0 µg/mL) and 100 µL of the secondary antibody (anti-mouse IgG-peroxidase, diluted 1:770 in TSB buffer, 1.0 µg/mL) per well for 1 h at room temperature. The binding was visualized as described before.

Optimization of MccJ25(RGDF) Production. For production optimization, M63 and M9 minimal medium (17.1 g/L $Na_2HPO_4 \cdot 12\ H_2O$, 3 g/L KH_2PO_4, 0.5 g/L NaCl, 1 g/L NH_4Cl, 1 mL/L $MgSO_4$ solution (2 M), 0.2 mL/L $CaCl_2$ solution (0.5 M), pH = 7.0, after autoclaving 10 mL/L glucose solution (40% w/v) and 0.2 mL/L vitamin mix (Supplementary Table S2) were added), containing 17 µg/mL chloramphenicol, were tested. For quantification, fermentations were performed in 500 mL M63 medium for 1 day at 37 °C, in 500 mL M9 medium for 1 day at 37 °C and in 100 mL M9 medium for 3 days at 37 °C. Supernatants of these fermentations were extracted as described above. The dried extracts were resuspended in 7.5 mL of 40% MeOH, cleared by centrifugation and then applied to purification by two rounds of preparative HPLC as described above. In this way, pure MccJ25(RGDF) could be obtained in yields of 4.6 mg/L (M63 medium, 1 day at 37 °C), 10.8 mg/L (M9 medium, 1 day at 37 °C) and 13.5 mg/L (M9 medium, 3 days at 37 °C) culture.

NMR Structure Elucidation of MccJ25(RGDF). The sample for the NMR measurements contained 7.0 mg of MccJ25(RGDF) in 250 µL of methanol-d_3. Spectra were recorded on a Bruker Avance 600 MHz spectrometer equipped with an inverse triple resonance $^1H/^{13}C/^{15}N$ probe with z-gradient. The sample was filled into a Wilmad 3 mm tube obtained from Rototec Spintec. Temperature effects on the structure were surveyed by recording ^1H spectra at variable temperatures. For sequential assignment, DQF-COSY, TOCSY, and NOESY spectra[9-12] were recorded at 298 and 303 K. TOCSY spectra were recorded with a mixing time of 140 ms. NOESY spectra were taken with mixing times of 100 and 300 ms. Solvent suppression was fulfilled by using excitation sculpting with gradients for DQF-COSY, TOCSY, and NOESY experiments.[13] 1D spectra were acquired with 65536 data points, while 2D spectra were collected using 4096 points in the F_2 dimension and 512 increments in the F_1 dimension. For 2D spectra 32 transients were used.

Relaxation delay was 3.0 s. Chemical shifts were referenced to solvent signal. All spectra were processed with Bruker TOPSPIN 3.1. NOE cross-peaks were further analyzed within the program Sparky.[14]

The assignment of ^{1}H chemical shifts was fulfilled by using ^{1}H – ^{1}H homonuclear correlation spectra.[15] Thus, a combination of DQF-COSY (Supplementary Figure S4) and NOESY (Supplementary Figure S5) produced sequential assignments (i.e. all αH and NH and their sequence in the backbone), and a combination of DQF-COSY and TOCSY (Supplementary Figure S6) resulted in assignments of the side chains (Supplementary Table S3). NOE cross-peaks observed in the 100 ms mixing time NOESY spectrum at 298 K were assigned and a list of peak volumes was obtained within Sparky. On the basis of the peak volumes of the available geminal protons of the glycine residues at positions 1, 2, 4, and 13, a manual distance calibration was carefully done with a reference distance of 1.78 Å.[16] The NOE peak volumes were thus converted into distances and used as distance upper limits for structure calculation within the program package CYANA 2.1.[17] Furthermore, an internal linkage for the isopeptide bond was realized by putting a distance constraint of 1.33 Å between the nitrogen atom of Gly1 and the Cδ of Glu8. In this way, 154 distance constraints were obtained; 58 for the backbone, 35 for long-range interactions and 96 for the side chains. Hence, there was an average of 7.0 distance constraints per residue. With the vicinal coupling constants $^{3}J_{HN-H}{}^{\alpha}$, which were measured directly from the amide proton signals in the 1D ^{1}H NMR spectra, torsion angles ϕ were calculated on the basis of the Karplus equation optimized for ϕ in peptides.[16, 18, 19] The conformations of the side chains were established carefully,[16, 19, 20] and only unambiguous ones were included: His5, Glu8, Phe10, Arg12, and Phe15 with $t^{2}g^{3}$, Val6 and Asp14 with $g^{2}t^{3}$, and Tyr9 and Ser[18] with $g^{2}g^{3}$. For $t^{2}g^{3}$, $g^{2}t^{3}$, and $g^{2}g^{3}$ conformations around the Cα-Cβ bond, the torsion angle χ^{1} was constrained in the range of $-60 \pm 15°$, $180 \pm 15°$, and $60 \pm 15°$, respectively.

The above mentioned constraints were used in the simulated annealing protocol for the CYANA 2.1 calculation. The calculation started with 50 random conformers and the resulting structures were engineered by the program package Sybyl 7.3 to include the covalent linkage between the nitrogen of Gly1 and Cδ of Glu8, followed by an energy minimization using the TRIPOS force field within Sybyl 7.3. Thus, on the basis of low energies and minimal violations of the experimental data, a family of 15 structures was chosen. These 15 energy-minimized conformers show an average root-mean-square deviation (RMSD) for the backbone of 0.06 Å and were kept to represent the solution structure of the grafted lasso peptide MccJ25(RGDF) (Supplementary Table S4). The quality of these structures was evaluated using the program PROCHECK.[21] The structure has been deposited to RCSB Protein Data Bank, with the accession code 2MMT.

Molecular Modelling. First, a manual docking of the MccJ25(RGDF) averaged NMR structure in the crystal structure of the αvβ3 integrin in complex with cilengetide (PDB code: 1L5G) was accomplished. Specifically, we superimposed the coordinates of the Cα carbons within the RGDF motif in MccJ25(RGDF) with the corresponding atoms in cilengetide, which was then removed from the complex. Subsequently, to mimic the typical interaction pattern of RGD integrin ligands,[22] the side chains of Arg12 and Asp14 in the lasso peptide were properly optimized to interact with (αv)-Asp218 and with the metal ion at the MIDAS, respectively. The obtained complex was then submitted to an over 100-ns MD simulation with NAMD 2.9,[23] using the *ff99SBildn* Amber force field parameters.[24, 25] Prior to the MD simulations, the co-crystallized Mn^{2+} ions in the receptor were replaced with the more physiologically relevant Mg^{2+} ions, which were also used in the biochemical assays. For these ions, we used the parameters recently developed by Allnér and co-workers,[26] which favour an octahedral coordination geometry. The parameters for the isopeptide bond between the terminal amine group of Gly1 and the carboxylic side chain of Glu8 in the lasso peptide were generated with Antechamber.[27] Initially, charges were computed using the restrained electrostatic potential (RESP) fitting procedure.[28] The ESP was first calculated by means of the Gaussian09 package[29] using a 6-31G* basis set at Hartree-Fock level of theory, and then the RESP charges were obtained by a two-stages fitting procedure using Antechamber.[27] Missing bonds, angles, torsion and improper torsion angle parameters were then generated using the same program. The complex was then solvated in a 10.0 Å layer cubic water box using the TIP3P water model parameters.[30] Neutrality was ensured by adding six further Mg^{2+} ions. A 12 Å cutoff (switched at 10.0 Å) was used for atom pair interactions. The long-range electrostatic interactions were computed by means of the particle mesh Ewald (PME) method using a 1.0 Å grid spacing in periodic boundary conditions. The RATTLE algorithm[31] was applied to constrain bonds involving hydrogen atoms, and thus an integration 2 fs time step interval could be used. The system was minimized and heated up to 300 K while putting harmonic constraints, which were gradually released along the thermalization process. The Mg^{2+} ions and their coordination shells were further equilibrated for 5 ns during which harmonic constraints of 0.5 kcal/mol were applied to these ions and the surrounding residues. Production runs were then performed under NPT conditions at 1 atm and 300 K.

Tube Formation Assays. The tube formation assay was performed as previously described.[32] Briefly, Matrigel™ (BD Discovery Labware, Bedford, MA) was filled into 'µ-slide Angiogenesis' chambers (IBIDI, Martinsried, Germany) and incubated at 37 °C. HUVECs were purchased from Promocell (Heidelberg, Germany). 12,000 HUVECs were added and stimulated with the respective

amount of each compound or left untreated as control reaction. After 16 h, images were taken on a Axiovert 200 microscope (Zeiss). The images were analyzed with the tube formation module of Wimasis (Munich, Germany). This software is programmed to recognize tubular structures and to analyze the area covered by tubes, tube length and network complexity (number of nodes). All experiments were done in triplicates (see Supplementary Table S5).

Cell Adhesion Assays. 12,000 HUVECs per well were pipetted into 96-well plates and treated with either cilengitide, MccJ25(RGD) or MccJ25(RGDF) at concentrations of 3, 10 or 30 μM. In a control reaction, the HUVECs were not further treated at all. After subsequent 15 min in a cell culture incubator, the plates were washed twice with warm PBS (37 °C). Then, the firmly adherent cells were fixed and stained with a crystal violet solution (0.5% crystal violet in 20% methanol). After washing and air drying, the bound dye was eluted from the cells with 0.1 M sodium citrate in 50% ethanol and absorbance was determined at 540 nm with a Sunrise plate reader (TECAN).

Supplementary Tables

Supplementary Table S1. Oligonucleotide primers for the generation of the MccJ25 variants via site-directed ligase-independent mutagenesis. SLIM overhangs are underlined. Introduced mutations are highlighted in bold.

name	sequence
mcjA_SLIM_FP	TTC TAT GGC TGA TAT TCT GAA AGA AGA ACT CTG
mcjA_SLIM_RP	CTC AGG CAC ATG TCC TGC ACC ACC
mcjA_SLIM_FRGD_Tail_FP	TAT TTT **TTC CGC GGC GAT** ACA CCT ATA TCT TTC TAT GGC TGA TAT TCT GAA AGA AGA ACT CTG
mcjA_SLIM_FRGD_Tail_RP	AGA TAT AGG TGT **ATC GCC GCG GAA** AAA ATA CTC AGG CAC ATG TCC TGC ACC ACC
mcjA_SLIM_RGDF_Tail_FP	TAT TTT GTG **CGC GGC GAT TTT** CCT ATA TCT TTC TAT GGC TGA TAT TCT GAA AGA AGA ACT CTG
mcjA_SLIM_RGDF_Tail_RP	AGA TAT AGG **AAA ATC GCC GCG** CAC AAA ATA CTC AGG CAC ATG TCC TGC ACC ACC

Supplementary Table S2. M9 Vitamin Mix.

component	amount
choline chloride	1.0 g
folic acid	1.0 g
pantothenic acid	1.0 g
nicotinamide	1.0 g
myo-inositol	2.0 g
pyridoxal hydrochloride	1.0 g
thiamine	1.0 g
riboflavin	0.1 g
disodium adenosine 5'-triphosphate	0.3 g
biotin	0.2 g
	ad 300 mL ddH_2O[a]

[a]10 M NaOH was slowly added to the resulting suspension until everything was dissolved (pH ~12). Afterward the solution was sterile filtered and stored at 4 °C until needed.

Supplementary Table S3. ^{1}H chemical shifts of MccJ25(RGDF) in methanol-d_3 at 298 K.

AA	NH	HA	HB	Others	AA	NH	HA	HB	Others
Gly1	7.952	4.229 3.544			Arg12	8.361	4.144	1.885 1.802	γCH_2: 1.625; 1.669 δCH_2: 3.203 NH: 7.410
Gly2	9.027	4.208 3.877			Gly13	8.492	4.071 3.541		
Ala3	8.543	4.655	1.307		Asp14	8.253	4.663	2.751 2.648	
Gly4	7.776	4.060 3.493			Phe15	8.212	4.936	2.971 3.203	$H_{2,6}$: 7.300; $H_{3,5}$: 7.300
His5	7.362	4.710	3.259 2.920	H_2: 8.710; H_4: 7.370	Pro16		4.451	1.891 1.748	γCH_2: 2.070; 1.957 δCH_2: 3.868; 3.569
Val6	8.703	4.687	1.746	γCH_3: 1.098; 0.848	Ile17	7.820	4.388	1.870	γCH_3: 0.877 γCH_2: 1.454 δCH_3: 1.121
Pro7		4.238	1.807 1.613	γCH_2: 2.200; 1.965 δCH_2: 4.130; 3.796	Ser18	7.415	4.405	4.048 3.860	
Glu8	8.288	4.347	1.764 1.652	γCH_2: 2.002; 1.858	Phe19	8.832	5.442	2.582 2.539	$H_{2,6}$: 6.863; $H_{3,5}$: 7.060
Tyr9	7.476	4.498	2.908 2.675	$H_{2,6}$: 6.926; $H_{3,5}$: 6.618	Tyr20	9.509	4.934	3.024 2.983	$H_{2,6}$: 6.966; $H_{3,5}$: 6.715
Phe10	8.093	4.795	2.941 2.764	$H_{2,6}$: 7.020; $H_{3,5}$: 7.150	Gly21	8.623	3.856		
Val11	8.218	4.237	2.082	γCH_3: 0.942; 0.942					

Supplementary Table S4. Structural statistics for the family of 15 structures selected to represent the solution structure of MccJ25(RGDF).

restraining constraints	constraint violations
total: 183	distance violations, >0.5 Å: 0
distance, i=j: 78	RMS deviations: 0.02 Å
distance, i-j=1: 41	dihedral violations, > 5°: 0
distance, i-j>1: 35	RMS deviation: 1.6°
dihedral: 29	average pairwise RMS deviation (Gly^2-Tyr^{20})
hydrogen bond: 0	backbone atoms: 0.06 Å
constraints/residue: 8.7	all heavy atoms: 0.40 Å

Supplementary Table S5a. Observed area covered by tubes in the tube formation assays.

compound	concentration	experiment 1	experiment 2	experiment 3	mean
none (control)	-	26	22	18	22.0 ± 3.3
cilengtide	3 μM	17	13	11	13.7 ± 2.5
	10 μM	17	12	12	13.7 ± 2.4
	30 μM	12	14	10	12.0 ± 1.6
MccJ25(RGD)	3 μM	17	18	13	16.0 ± 2.2
	10 μM	16	15	11	14.0 ± 2.2
	30 μM	13	14	14	13.7 ± 0.5
MccJ25(RGDF)	3 μM	18	15	13	15.3 ± 2.1
	10 μM	14	14	10	12.7 ± 1.9
	30 μM	14	13	9	12.0 ± 2.2

Supplementary Table S5b. Observed tube length in the tube formation assays.

compound	concentration	experiment 1	experiment 2	experiment 3	mean
none (control)	-	12883	10852	9116	10950 ± 1539
cilengtide	3 μM	9241	7443	6321	7668 ± 1203
	10 μM	8463	6943	5898	7101 ± 1053
	30 μM	8542	8645	5284	7490 ± 1561
MccJ25(RGD)	3 μM	9419	10522	7080	9007 ± 1435
	10 μM	9285	7575	6447	7769 ± 1167
	30 μM	7600	7438	6023	7020 ± 708
MccJ25(RGDF)	3 μM	10167	8643	6402	8404 ± 1546
	10 μM	8513	7214	5927	7218 ± 1056
	30 μM	7209	7467	5280	6652 ± 976

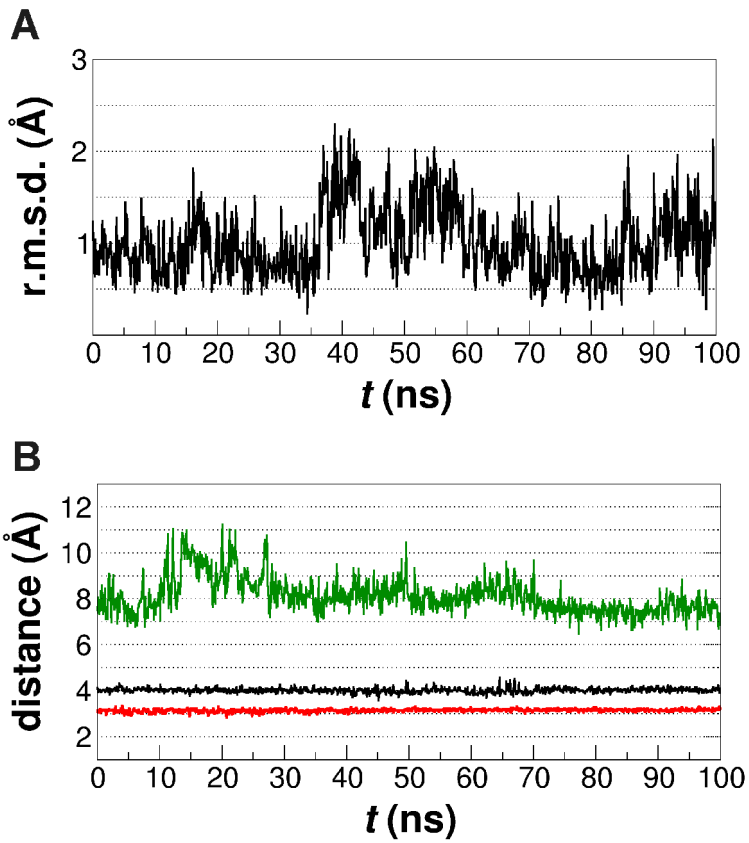
Supplementary Table S5c. Number of branching points observed in the tube formation assays.

compound	concentration	experiment 1	experiment 2	experiment 3	mean
none (control)	-	68	49	43	53.3 ± 10.7
cilengtide	3 μM	33	24	25	27.3 ± 4.0
	10 μM	29	19	20	22.7 ± 4.5
	30 μM	28	33	11	24.0 ± 9.4
MccJ25(RGD)	3 μM	42	48	25	38.3 ± 9.7
	10 μM	46	33	17	32.0 ± 11.9
	30 μM	30	31	14	25.0 ± 7.8
MccJ25(RGDF)	3 μM	48	43	21	37.3 ± 11.7
	10 μM	38	24	18	26.7 ± 8.4
	30 μM	27	27	20	24.7 ± 3.3

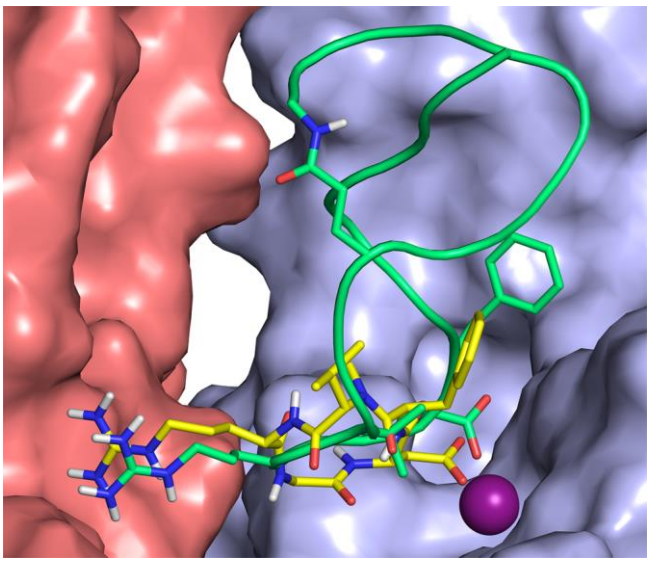
Supplementary Table S5d. Inhibition of cell adhesion of HUVECs relative to a control reaction.

compound	concentration	experiment 1	experiment 2	experiment 3	mean
none (control)	-	99.4	92.8	107.8	100.0 ± 6.1
cilengtide	3 μM	53.3	55.6	61.7	56.9 ± 3.5
	10 μM	50.6	51.7	54.4	52.2 ± 1.6
	30 μM	53.3	50.6	50.6	51.5 ± 1.3
MccJ25(RGD)	3 μM	76.1	77.8	75.6	76.5 ± 0.9
	10 μM	58.9	65.0	62.8	62.2 ± 2.5
	30 μM	60.0	53.3	61.7	58.3 ± 3.6
MccJ25(RGDF)	3 μM	96.7	83.9	90.6	90.4 ± 5.2
	10 μM	72.2	70.6	75.0	72.6 ± 1.8
	30 μM	65.0	62.2	71.7	66.3 ± 4.0

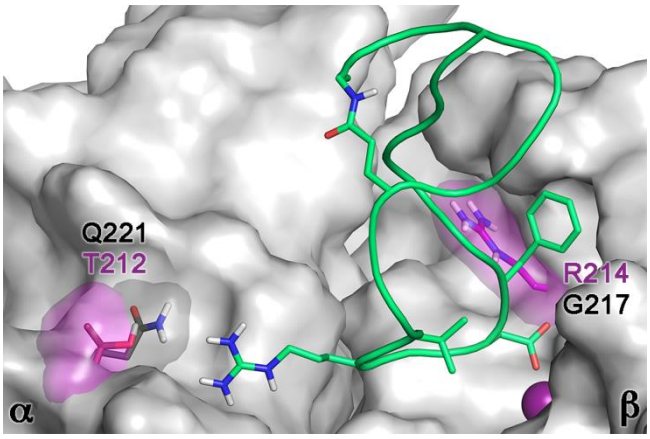
Supplementary Figures



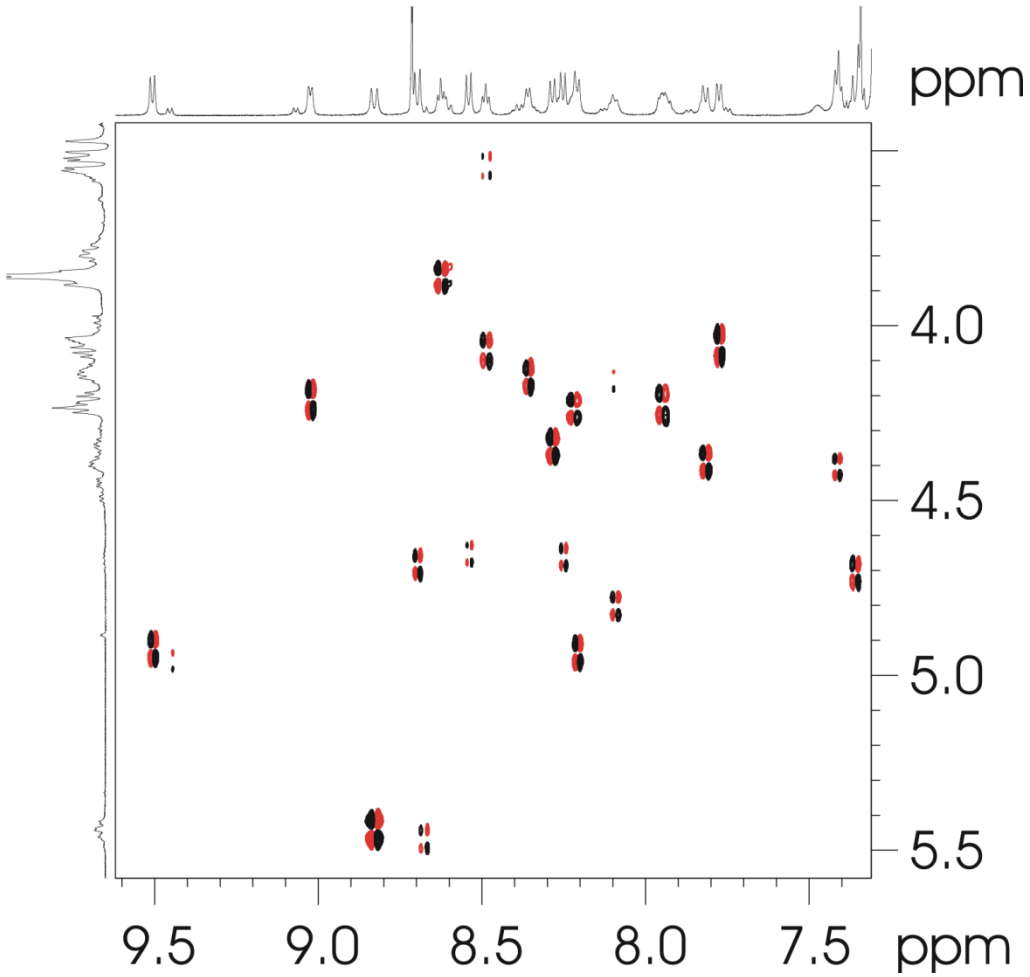
Supplementary Figure S1. (A) The averaged rmsd plot of the Cα carbons within the RGDF motif in MccJ25(RGDF) during the production MD run. Prior to the rmsd calculations, trajectory frames were aligned using the coordinates of the Cα carbons of the secondary structural elements of the αvβ3 integrin. (B) Time evolution of the distances between representative atoms of the amino acids involved in the key MccJ25(RGDF)/αvβ3 interactions described in the main text: Asp14 Cγ–Mg^{2+} ion at the MIDAS (red plot); Arg12 Cζ–(αv)-Asp218 Cγ (black plot); Arg12 Cζ–(αv)-Thr212 Cβ (green plot).



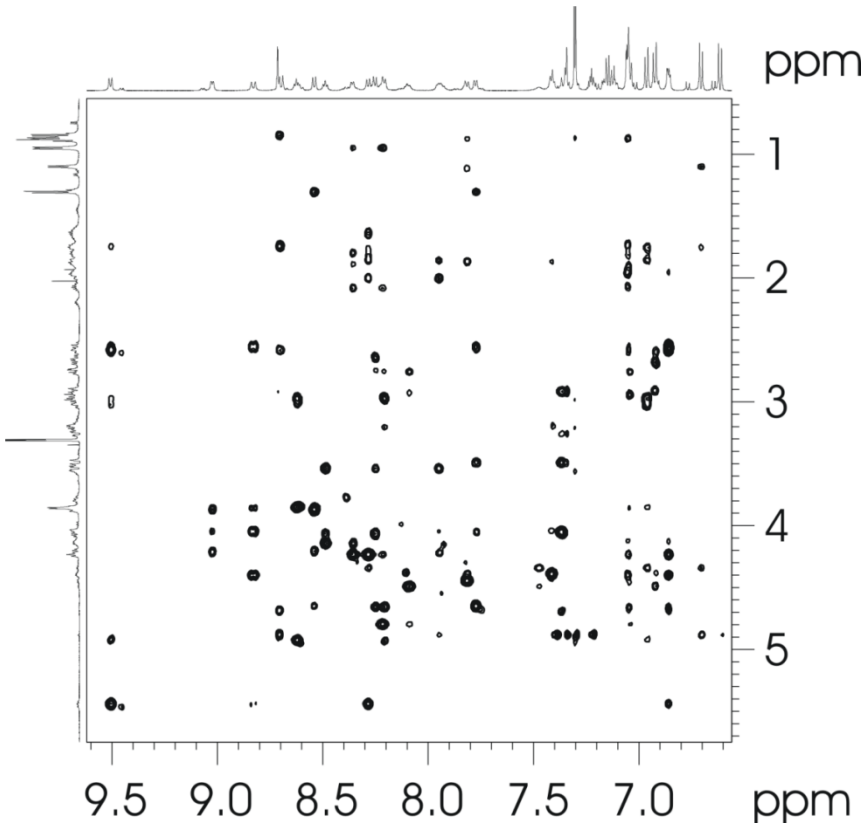
Supplementary Figure S2. Comparison between the binding poses of MccJ25(RGDF) and cilengetide at the αvβ3 integrin. The lasso peptide is shown as green cartoon and sticks, while cilengetide is depicted as yellow sticks. The αv and β3 receptor subunits are represented as light red and blue surface, respectively. The metal ion at the MIDAS is shown as a purple sphere. The crystal structure of the αvβ3 integrin employed in the MD simulation was obtained from the protein data bank (PDB code 1L5G).



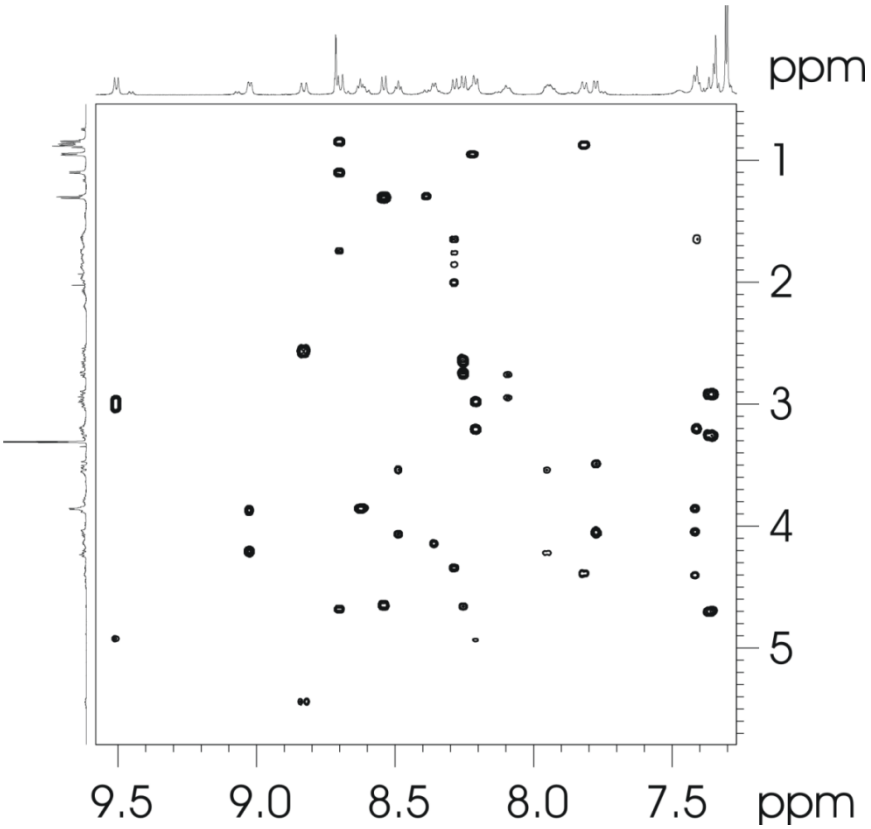
Supplementary Figure S3. Superposition between the α5β1 receptor and the αvβ3 integrin in complex with MccJ25(RGDF). The lasso peptide is shown as green cartoon and sticks. The α5β1 and αvβ3 (PDB code 1L5G) receptors (only key substitutions) are represented as gray and transparent magenta Connolly surfaces, respectively. In both receptors, residues important for selectivity are highlighted as sticks. The metal ion at the MIDAS is shown as a purple sphere.



Supplementary Figure S4. Fingerprint region of the DQF-COSY spectrum of MccJ25(RGDF) in methanol-d_3 at 298 K.



Supplementary Figure S5. Section of the NOESY spectrum of MccJ25(RGDF) in methanol-d_3 at 298 K and with a mixing time of 300 ms.



Supplementary Figure S6. Section of the TOCSY spectrum of MccJ25(RGDF) in methanol-d_3 at 298 K and with a mixing time of 140 ms.

References

(1) Solbiati, J. O.; Ciaccio, M.; Farias, R. N.; Gonzalez-Pastor, J. E.; Moreno, F.; Salomon, R. A. Sequence analysis of the four plasmid genes required to produce the circular peptide antibiotic microcin J25. *J. Bacteriol.* **1999,** *181*, 2659-2662.

(2) Chiu, J.; March, P. E.; Lee, R.; Tillett, D. Site-directed, Ligase-Independent Mutagenesis (SLIM): a single-tube methodology approaching 100% efficiency in 4 h. *Nucleic Acids Res.* **2004,** *32*, e174.

(3) Chiu, J.; Tillett, D.; Dawes, I. W.; March, P. E. Site-directed, Ligase-Independent Mutagenesis (SLIM) for highly efficient mutagenesis of plasmids greater than 8kb. *J. Microbiol. Methods* **2008,** *73*, 195-198.

(4) Knappe, T. A.; Manzenrieder, F.; Mas-Moruno, C.; Linne, U.; Sasse, F.; Kessler, H.; Xie, X.; Marahiel, M. A. Introducing lasso peptides as molecular scaffolds for drug design: engineering of an integrin antagonist. *Angew. Chem. Int. Ed.* **2011,** *50*, 8714-8717.

(5) Marugan, J. J.; Manthey, C.; Anaclerio, B.; Lafrance, L.; Lu, T.; Markotan, T.; Leonard, K. A.; Crysler, C.; Eisennagel, S.; Dasgupta, M.; Tomczuk, B. Design, synthesis, and biological evaluation of novel potent and selective alphavbeta3/alphavbeta5 integrin dual inhibitors with improved bioavailability. Selection of the molecular core. *J. Med. Chem.* **2005,** *48*, 926-934.

(6) Stragies, R.; Osterkamp, F.; Zischinsky, G.; Vossmeyer, D.; Kalkhof, H.; Reimer, U.; Zahn, G. Design and synthesis of a new class of selective integrin alpha5beta1 antagonists. *J. Med. Chem.* **2007,** *50*, 3786-3794.

(7) Hartman, G. D.; Egbertson, M. S.; Halczenko, W.; Laswell, W. L.; Duggan, M. E.; Smith, R. L.; Naylor, A. M.; Manno, P. D.; Lynch, R. J.; Zhang, G.; et al. Non-peptide fibrinogen receptor antagonists. 1. Discovery and design of exosite inhibitors. *J. Med. Chem.* **1992,** *35*, 4640-4642.

(8) Dechantsreiter, M. A.; Planker, E.; Matha, B.; Lohof, E.; Holzemann, G.; Jonczyk, A.; Goodman, S. L.; Kessler, H. N-Methylated cyclic RGD peptides as highly active and selective alpha(V)beta(3) integrin antagonists. *J. Med. Chem.* **1999,** *42*, 3033-3040.

(9) Jeener, J.; Meier, B. H.; Bachmann, P.; Ernst, R. R. Investigation of Exchange Processes by 2-Dimensional NMR-Spectroscopy. *J. Chem. Phys.* **1979,** *71*, 4546-4553.

(10) Rance, M.; Sorensen, O. W.; Bodenhausen, G.; Wagner, G.; Ernst, R. R.; Wüthrich, K. Improved spectral resolution in cosy 1H NMR spectra of proteins via double quantum filtering. *Biochem. Biophys. Res. Commun.* **1983,** *117*, 479-485.

(11) Bax, A.; Davis, D. G. Mlev-17-Based Two-Dimensional Homonuclear Magnetization Transfer Spectroscopy. *J. Magn. Reson.* **1985,** *65*, 355-360.

(12) Marion, D.; Ikura, M.; Tschudin, R.; Bax, A. Rapid Recording of 2d NMR-Spectra without Phase Cycling - Application to the Study of Hydrogen-Exchange in Proteins. *J. Magn. Reson.* **1989,** *85*, 393-399.

(13) Hwang, T. L.; Shaka, A. J. Water Suppression That Works - Excitation Sculpting Using Arbitrary Wave-Forms and Pulsed-Field Gradients. *J. Magn. Reson. A* **1995,** *112*, 275-279.

(14) Goddard, T. D.; Kneller, D. J. *Sparky 3*, University of California, San Fransisco: 2007.

(15) Wüthrich, K. *NMR of proteins and nucleic acids*. Wiley: New York, 1986.

(16) Beck, J. G. F., A. O.; Kessler, H. *NMR of Biomolecules*. Wiley-VCH: Weinheim, 2012.

(17) Herrmann, T.; Güntert, P.; Wüthrich, K. Protein NMR structure determination with automated NOE assignment using the new software CANDID and the torsion angle dynamics algorithm DYANA. *J. Mol. Biol.* **2002,** *319*, 209-227.

(18) Pardi, A.; Billeter, M.; Wüthrich, K. Calibration of the angular dependence of the amide proton-C alpha proton

coupling constants, 3JHN alpha, in a globular protein. Use of 3JHN alpha for identification of helical secondary structure. *J. Mol. Biol.* **1984,** *180*, 741-751.

(19) Wagner, G. NMR Investigations of Protein-Structure. *Progress in Nuclear Magnetic Resonance Spectroscopy* **1990,** *22*, 101-139.

(20) Clore, G. M.; Gronenborn, A. M. Determination of three-dimensional structures of proteins and nucleic acids in solution by nuclear magnetic resonance spectroscopy. *Crit. Rev. Biochem. Mol. Biol.* **1989,** *24*, 479-564.

(21) Laskowski, R. A.; Rullmannn, J. A.; MacArthur, M. W.; Kaptein, R.; Thornton, J. M. AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. *J. Biomol. NMR* **1996,** *8*, 477-486.

(22) Mas-Moruno, C.; Rechenmacher, F.; Kessler, H. Cilengitide: the first anti-angiogenic small molecule drug candidate design, synthesis and clinical evaluation. *Anticancer Agents Med. Chem.* **2010,** *10*, 753-768.

(23) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kale, L.; Schulten, K. Scalable molecular dynamics with NAMD. *J. Comput. Chem.* **2005,** *26*, 1781-1802.

(24) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. *J. Am. Chem. Soc.* **1995,** *117*, 5179-5197.

(25) Lindorff-Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Klepeis, J. L.; Dror, R. O.; Shaw, D. E. Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins* **2010,** *78*, 1950-1958.

(26) Allnér, O.; Nilsson, L.; Villa, A. Magnesium Ion–Water Coordination and Exchange in Biomolecular Simulations. *J. Chem. Theo. Comp.* **2012,** *8*, 1493-1502.

(27) Wang, J.; Wang, W.; Kollman, P. A.; Case, D. A. Automatic atom type and bond type perception in molecular mechanical calculations. *J. Mol. Graph. Model* **2006,** *25*, 247-260.

(28) Bayly, C. I.; Cieplak, P.; Cornell, W.; Kollman, P. A. A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model. *J. Phys. Chem.* **1993,** *97*, 10269-10280.

(29) *Gaussian 09*, Gaussian 09, Revision D.01

Gaussian, Inc.: Wallingford CT, 2009.

(30) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids. *J. Am. Chem. Soc.* **1996,** *118*, 11225-11236.

(31) Andersen, H. C. Rattle: A "velocity" version of the shake algorithm for molecular dynamics calculations. *J. Comp. Phys.* **1983,** *52*, 24-34.

(32) Rothmeier, A. S.; Ischenko, I.; Joore, J.; Garczarczyk, D.; Furst, R.; Bruns, C. J.; Vollmar, A. M.; Zahler, S. Investigation of the marine compound spongistatin 1 links the inhibition of PKCalpha translocation to nonmitotic effects of tubulin antagonism in angiogenesis. *FASEB J.* **2009,** *23*, 1127-1137.